Le texte dans son contexte : Statistique

Michel Armatte (Alexandre Koyré, histoire des sciences)

Workshop : *Un siècle de Fisher*

Analyse de variance et études d'héritabilité depuis *(1918)*

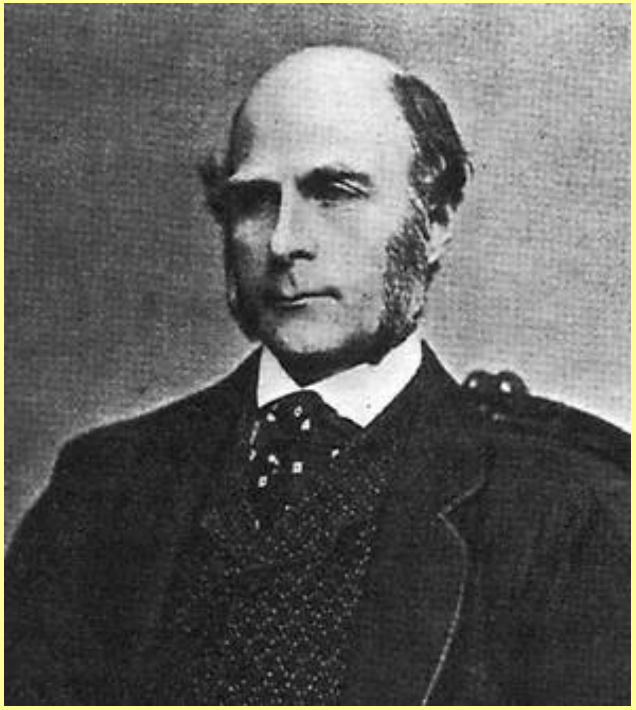
The correlation between relatives on the Supposition of Mendelian Inheritance

Fisher et la biométrie

- **Biométrie**
- 1. corrélation et régression chez Galton
- 2. contingence et corrélation chez Pearson
- 3. Régression et moindres carrés chez Yule
- 4. Biométrie et Génétique
- 5. Détour économétrique

- **Fisher**
- 6. Eléments de biographie
- 7. L'œuvre statistique
- 8. Fisher 1918 (CP9)
  - Découpage et contenu
  - Contexte de publication
- 9. Quelques évaluations
- 10 Conclusion

1. corrélation et régression chez Galton



Francis Galton (1822-1911), cousin de Darwin

- Famille quaker victorienne de 9 enfants. Sait lire et compter à 4 ans
- Etudes légères en math et biologie
- Voyages en Afrique : *The Art of Travel (1855).* Entrée à la *Royal Geographical Society.*
- *Principales inventions : anticyclone, carte météo isobars et isochroniques, sac de couchage, couper un cake, finger prints, portrait robot, coups de pinceaux…*
- Régression et corrélation
  - Perte de la foi : *J'étais misérablement écrasé par le poids du vieil "argument from design"*
  - Nouveau programme de recherche : l'hérédité et ses mécanismes (*transfusion, stirpes, enquêtes auprès des juges…)*

L'eugénisme

- 1865 : « *Si la vingtième partie des coûts et des peines qui sont dépensées dans l'amélioration de la reproduction des chevaux et du bétail étaient dépensés dans des mesures en faveur de l'amélioration de la race humaine, quelle galaxie de génies n'aurions nous pas créé!* »
- 1883 : *l'eugénique est la science qui traite de toutes les influences qui améliorent les qualités d'une race*
- l'eugénique positive vise à améliorer le taux de reproduction de la méritocratie, c'est à dire de celle qui est le plus « capable » et qui se sent en danger face à l'aristocratie, la ploutocratie, ou la classe laborieuse; l'eugénique négative vise à contrôler ce débordement par la limitation du taux de reproduction de cette frange inférieure de la population caractérisée comme inapte physiquement ou mentalement
- L'amélioration de la race humaine est donc à la fois un programme scientifique et un programme politique. Le premier tourne autour des lois de l'hérédité et de leur rôle dans le système variation-sélection. Le second comprend effectivement deux volets: l'eugénique positive vise à améliorer le taux de reproduction et la domination de cette classe méritante et méritocratique qui se sent parfois en danger face à la montée des puissances de l'argent ou face à l'explosion plus dangereuse des classes laborieuses
- Eugenic Laboratory, Eugenic Society, Eugenic Review

La loi de Laplace-Gauss était pour Quetelet le signe de l'homogénéité autour de la moyenne (le type)

La loi de "déviation par rapport à la moyenne, baptisée "loi normale" en 1889, est pour Galton le symbole de la variabilité et de l'hétérogénéité et un outil d'interclassement et de ségrégation

Je me propose de montrer dans ce livre que les moyens (abilities) naturel d'un homme dérivent par hérédité exactement comme la forme et les caractères physiques de tout être organisé*"Je propose de ranger les hommes en fonction de leurs capacités naturelles, en les mettant dans des classes séparées par des degrés égaux de mérite, et de montrer le nombre relatif d'individus inclus dans les différentes classes".*

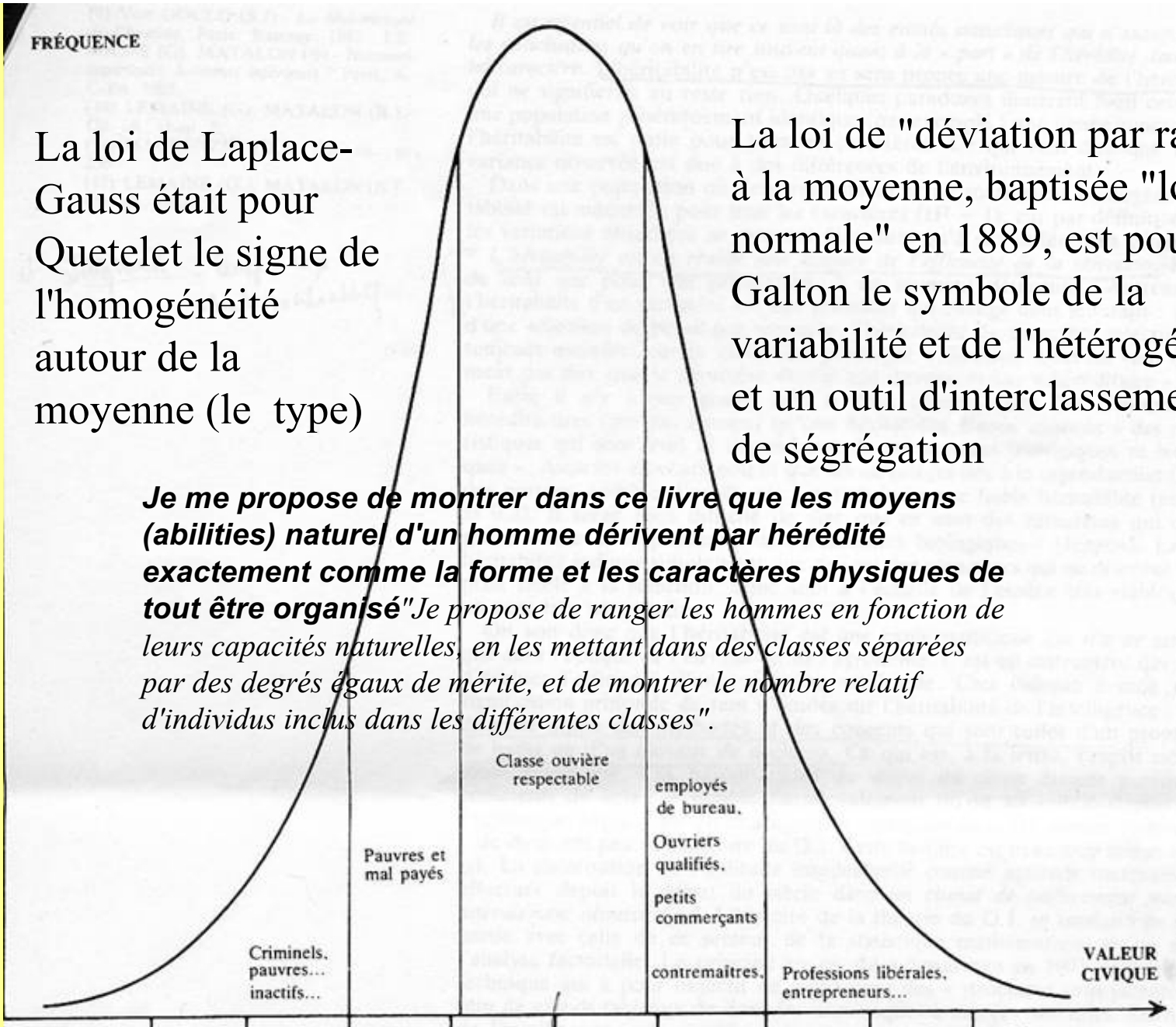


Diagramme par lequel Galton représente la distribution de la « valeur civique » dans la population londonnienne. La courbe est bien entendu purement théorique. Les classes « t, u, v » sont qualifiées comme « indésirables » ; T, U, V renvoient à la fraction de la population dont il faut au contraire encourager la fécondité. (F. GALTON, *Essays in Eugenics*, London, 1909, p. 11 sq).

La Quincunx ou machine de Galton

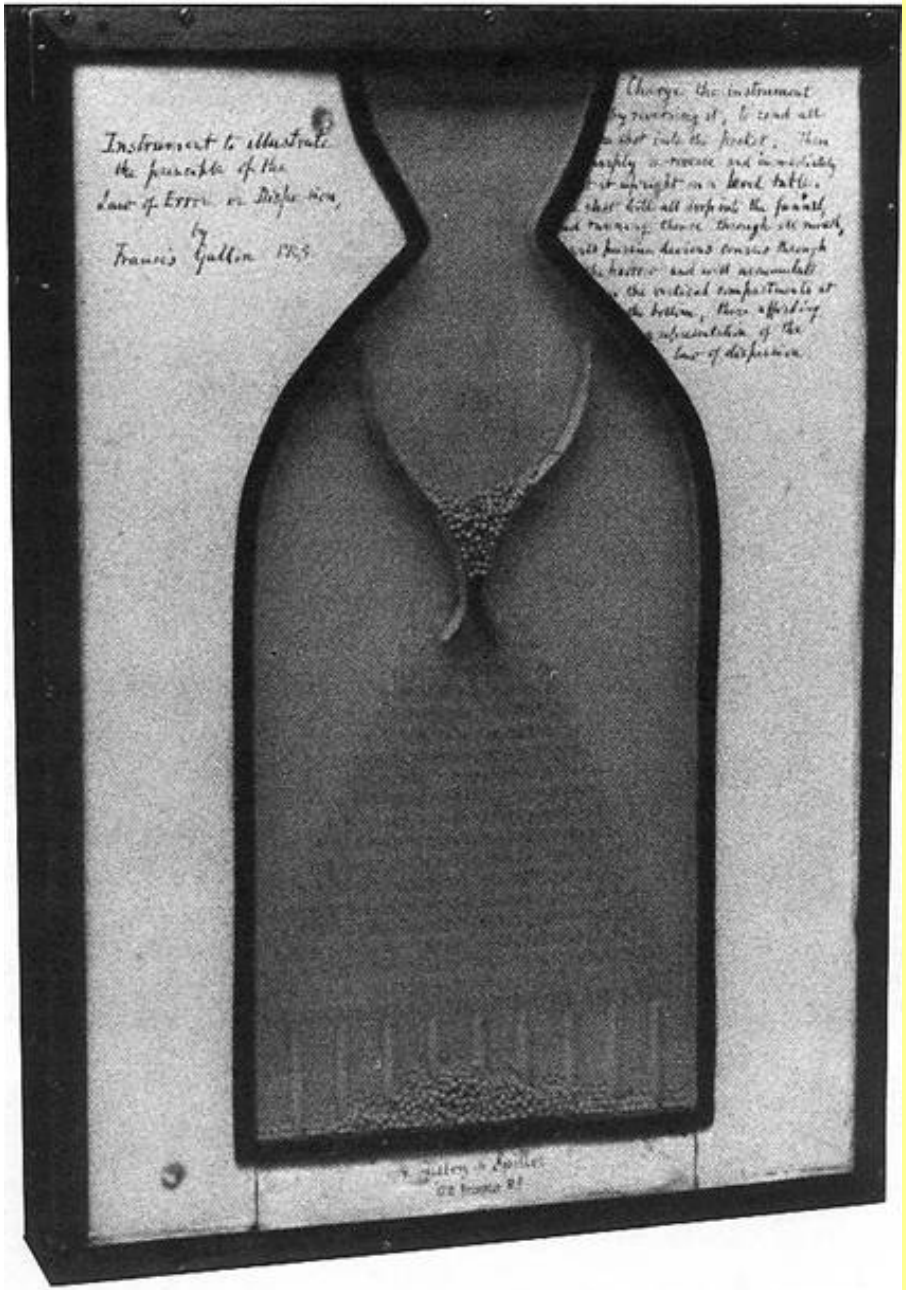
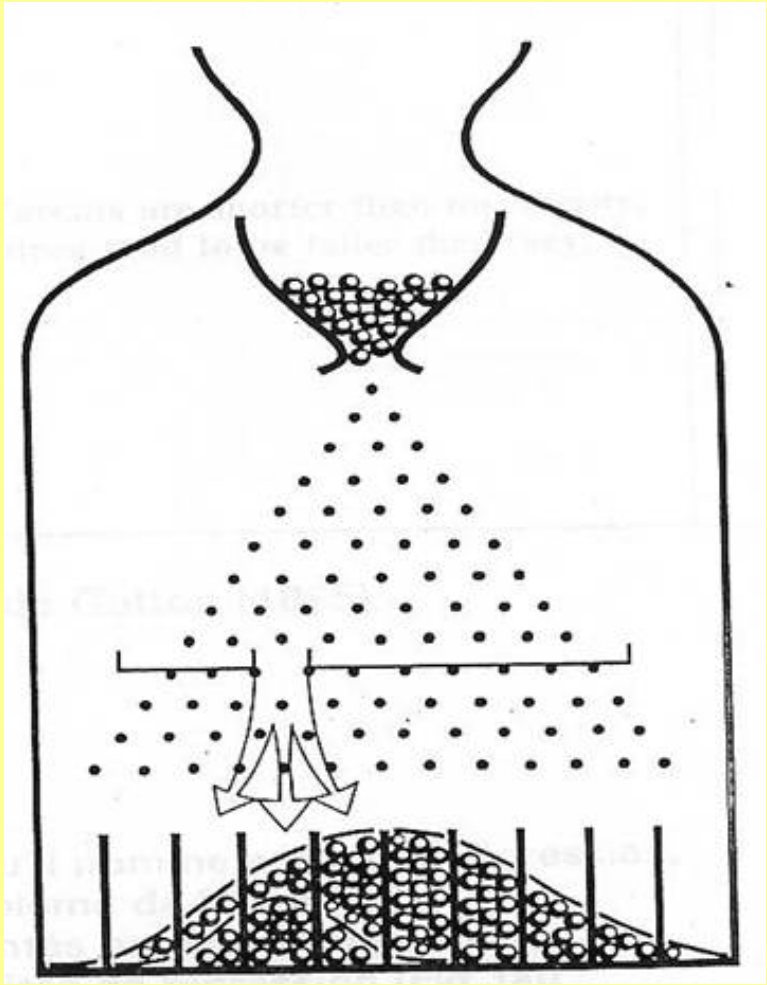
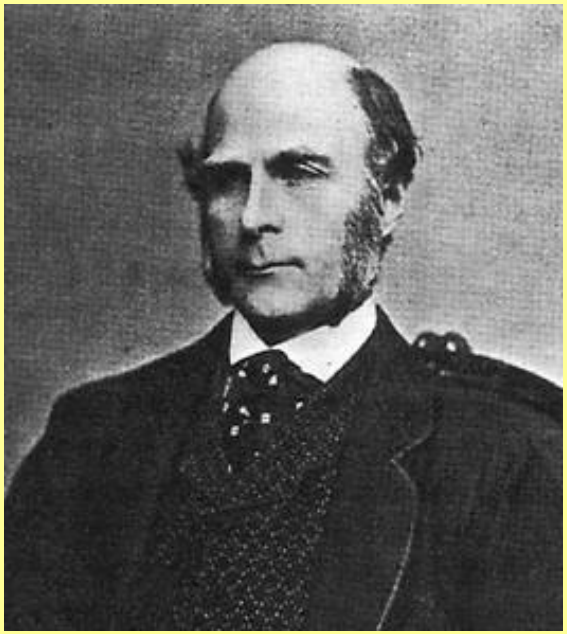


Figure 8.4. The original quincunx, apparently made for Galton in 1873 by Tisley & Spiller. Although it once had an opening at the top through which the shot could be poured, the top is now sealed with the shot inside. The glass has become cloudy with lead dust over the years. The caption, in Galton's handwriting, reads:

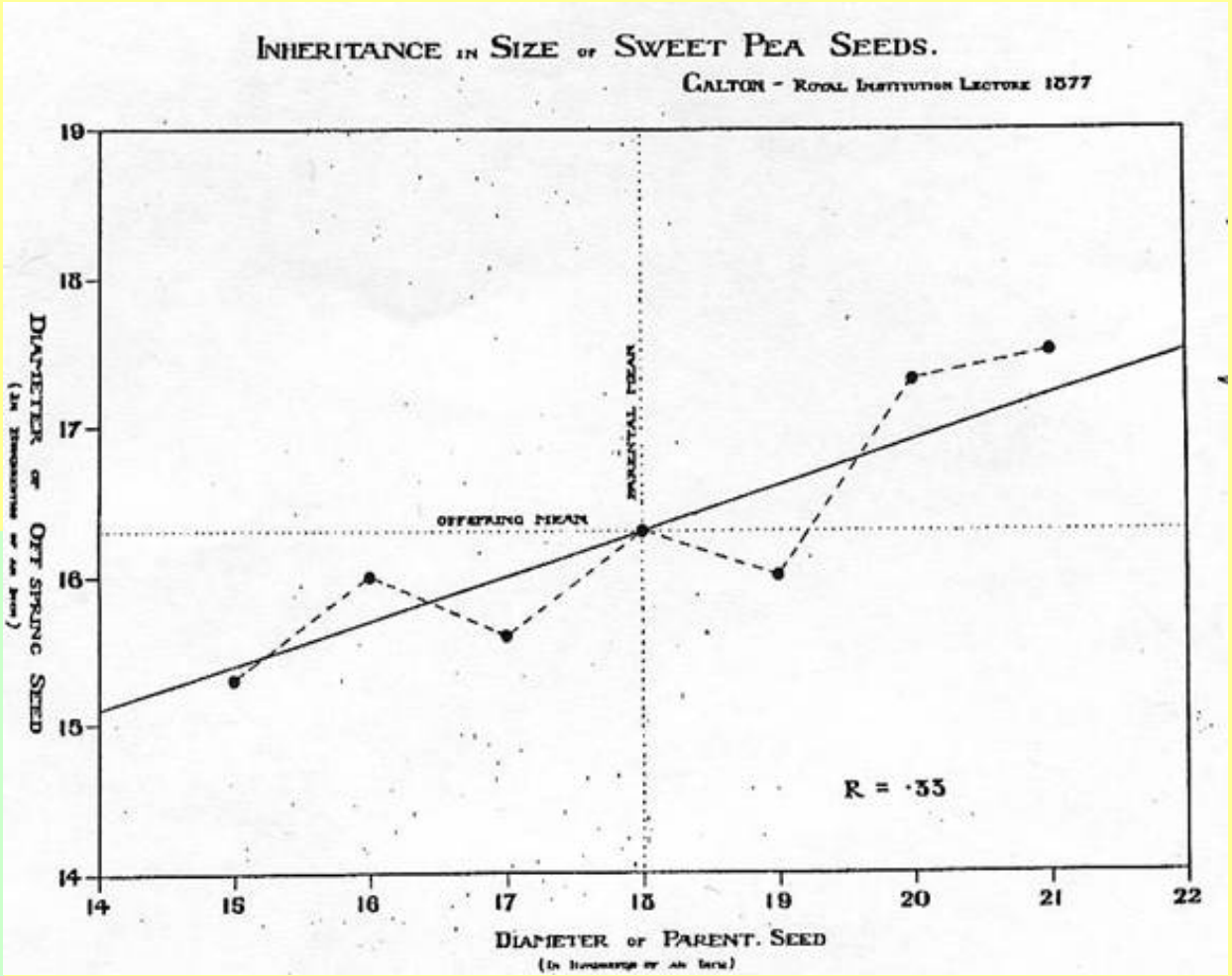


Cette machine permet de simuler une loi normale par convergence de loi binomiale : chaque petite bille en tombant sur un clou a une chance sur deux de passer à gauche ou à droite de ce clou, ce qui simule une loi de Bernoulli. La traversée des n rangées de clous placées en quinconces (quincunx) est analogue à la répétition indépendante de n épreuves de Bernoulli pour cette bille. La déviation finale de la bille suit donc une loi Binomiale (n,1/2) dont l'approximation est normale

F. Galton : les lois de la Réversion



- Quelles sont les lois de l'hérédité?
- Première (1875) expérience de Galton avec les pois de senteur : les graines mères donnent des graines filles de taille intermédiaire entre mère et type.



la Régression (1884) Taille des parents et taille des enfants adultes

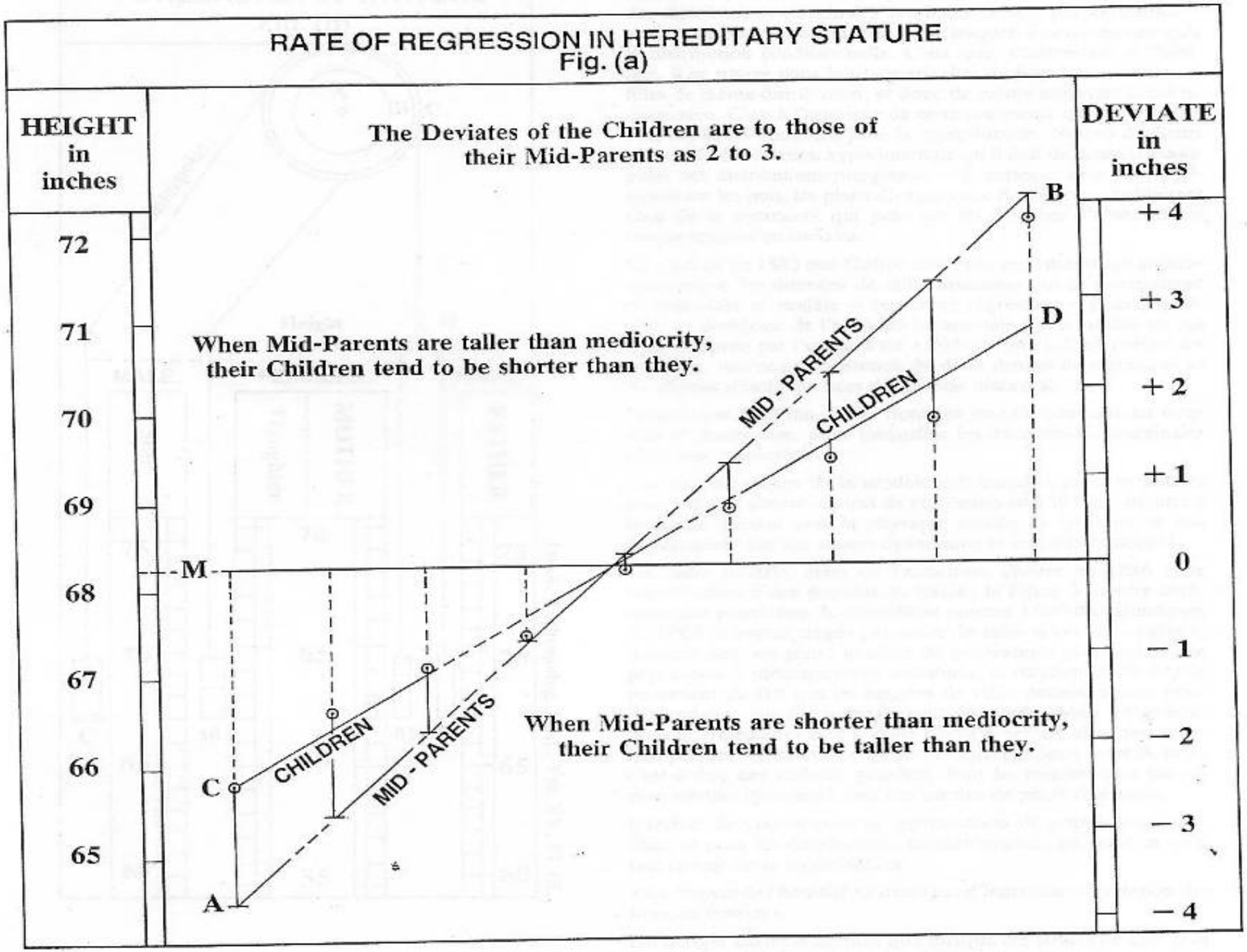
Table III (R.F.F. Data).

Number of Adult Children of various Statures born of 205 Mid-parents of various Statures.
(All Female Heights have been multiplied by 1·08).

Height of the mid-parents in inches.	Heights of the adult children.														Total number of		Medians.
	Below	62·2	63·2	64·2	65·2	66·2	67·2	68·2	69·2	70·2	71·2	72·2	73·2	Above.	Adult children.	Mid-parents.	
Above	..	..	..	..	..	..	..	..	..	..	..	1	3	..	4	5	
72·5....	..	..	..	..	..	..	..	1	2	1	2	7	2	4	19	6	72·2
71·5....	..	..	..	..	1	3	4	3	5	10	4	9	2	2	43	11	69·9
70·5....	1	..	1	..	1	1	3	12	18	14	7	4	3	3	68	22	69·5
69·5....	..	..	1	16	4	17	27	20	33	25	20	11	4	5	183	41	68·9
68·5....	1	..	7	11	16	25	31	34	48	21	18	4	3	..	219	49	68·2
67·5...	..	3	5	14	15	36	38	28	38	19	11	4	..	..	211	33	67·6
66·5....	..	3	3	5	2	17	17	14	13	4	..	..	..	..	78	20	67·2
65·5....	1	..	0	5	7	11	11	7	7	5	2	1	..	..	66	12	66·7
64·5....	1	1	4	4	1	5	5	..	2	..	..	..	..	..	23	5	65·8
Below	1	..	2	4	1	2	2	1	1	..	..	..	..	..	14	1	
Totals	5	7	32	59	48	117	138	120	167	99	64	41	17	14	928	205	
Medians ..	..	..	66·3	67·8	67·9	67·7	67·9	68·3	68·5	69·0	69·0	70·0					

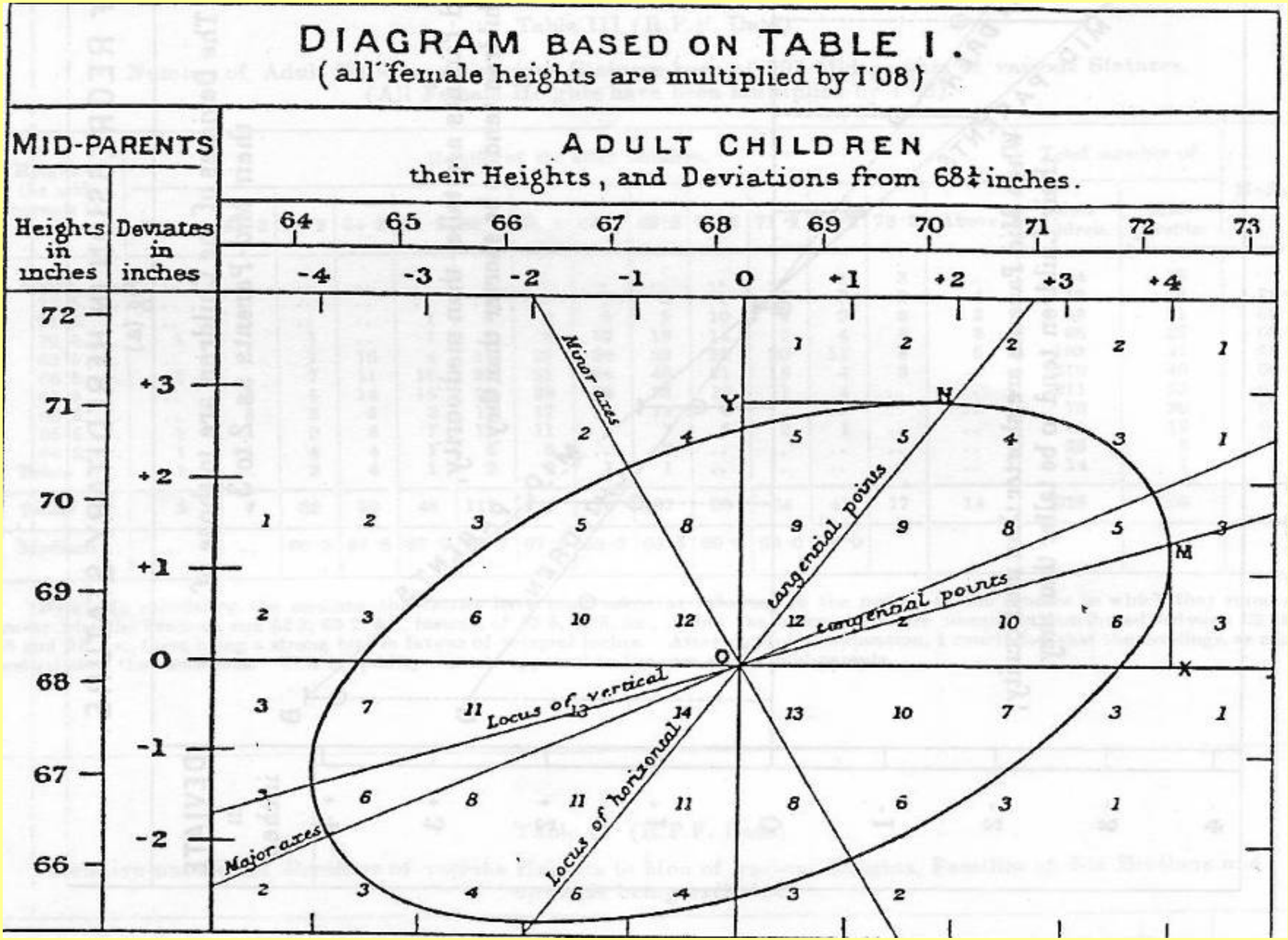
Note.—In calculating the medians, the entries have been taken as referring to the middle of the squares in which they stand. The reason why the headings run 62·2, 63·2, &c., instead of 62·5, 63·5, &c., is that the observations are unequally distributed between 62 and 63, 63 and 64, &c., there being a strong bias in favour of integral inches. After careful consideration, I concluded that the headings, as adopted, best satisfied the conditions. This inequality was not apparent in the case of the mid-parents.

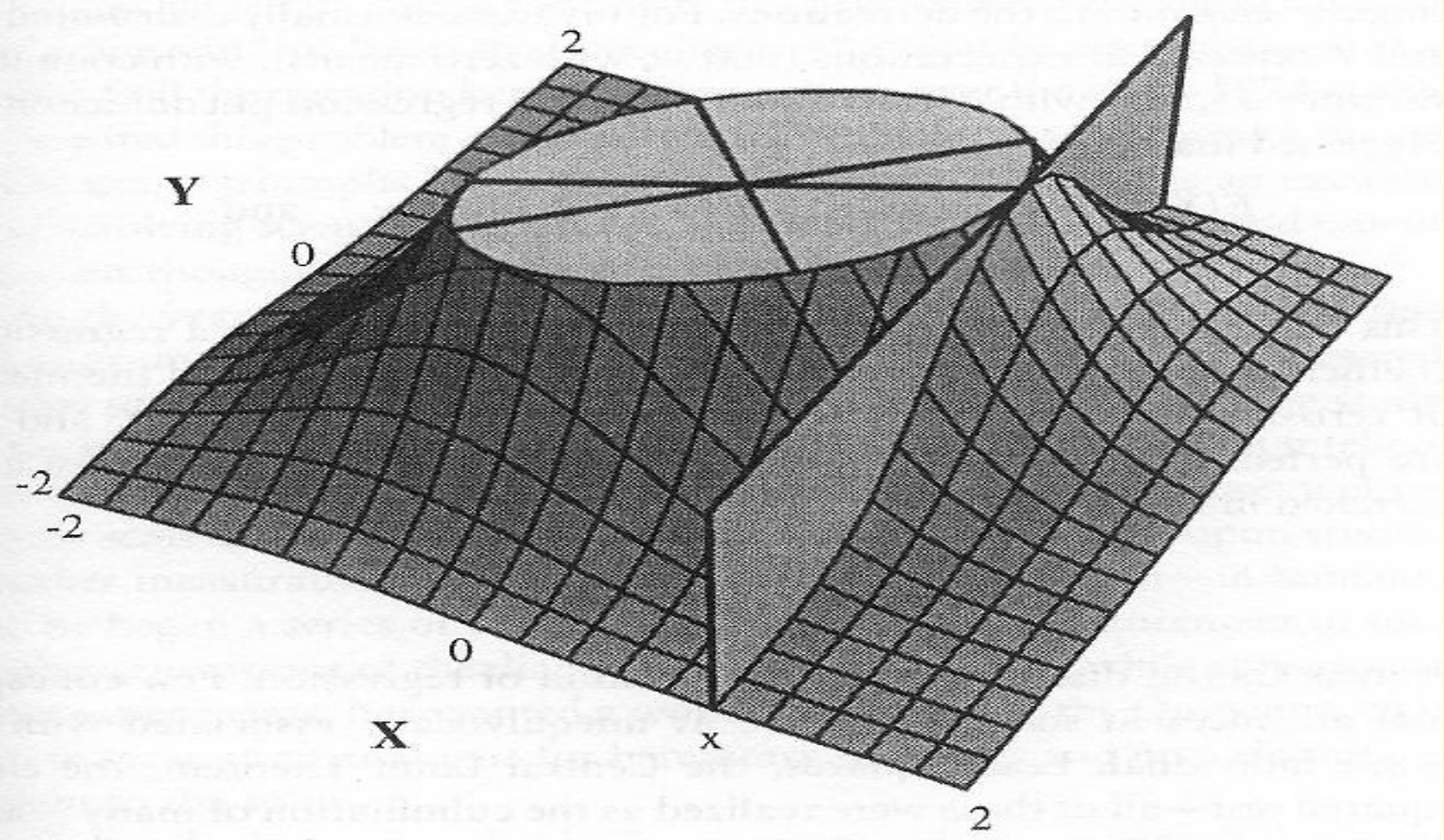
: "*Monsieur F. Galton offre 500£ de prix aux citoyens britanniques résidents dans le Royaume-Uni qui pourront lui fournir avant le 15 mai 1884 le meilleur extrait de leur propre fiche de famille"* concernant taille, couleur des yeux, tempérament, dispositions artistiques, maladies, caractéristiques du conjoint et descendance. Le questionnaire est adressé à des médecins et hommes de loi, et fournit des données (*RFF data*) concernant les tailles de 205 couples de parents et 928 enfants. L'année suivante, Galton saisit l'occasion d'une exposition internationale de médecine au musée des sciences de South Kensington pour installer à ses frais un laboratoire anthropométrique qui survivra encore 8 années à l'exposition : durant celle-ci, 9337 personnes viendront pour un droit d'entrée de 3 shillings, se faire mesurer par trois opérateurs sous toutes les coutures : poids, taille, capacité respiratoire, force, pouls, vue, ouïe.



d'après la planche IX de Galton (1885)

c. Ellipse d'équiprobabilité




2
Y
0
-2
-2
0
X
x
2

La Corrélation

$$[1] \quad E(y/x=x_i)=\bar{y}_i=ax_i+b$$

$$V(\bar{y}_i)=a^2V(x_i)=r^2V(y)\Rightarrow a=r\frac{\sigma_y}{\sigma_x}$$

$$[1] \text{ devient } [2]: \quad \frac{\bar{y}_i-\bar{y}}{\sigma_y}=r\frac{x_i-\bar{x}}{\sigma_x}$$

$$[1'] \quad E(x/y=y_i)=\bar{x}_i=cx_i+d$$

$$V(\bar{x}_i)=c^2V(y_i)=r^2V(x)\Rightarrow c=r\frac{\sigma_x}{\sigma_y}$$

$$[1'] \text{ devient } [2']: \quad \frac{\bar{x}_i-\bar{x}}{\sigma_x}=r\frac{y_i-\bar{y}}{\sigma_y}$$

- 1886 : régression de x (taille fils) en y(taille père) = 1/3
- r est la pente de la régression en données centrées réduites
- r² est le rapport variance des moyennes/variance totale
- 1888 : r est une mesure symétrique de corrélation biologique

Galton a l'idée de transporter son dispositif d'analyse des tables d'hérédité à n'importe quel tableau à double entrée formé par une distribution conjointe de deux variables statistiques. Il retrouve la notion de *co-relation* connue depuis longtemps en biologie. le coefficient de corrélation est une mesure symétrique de la liaison entre x et y qui la résume à lui seul .

2. Contingence et corrélation chez Karl Pearson (1857-1936)

- Fils d'avocat. Etudes de math à Londres et Cambridge, puis de droit

- Voyages en Allemagne : Karl Marx, le socialisme, la libre pensée, lae féminisme, l'impérialisme,l'eugénisme.

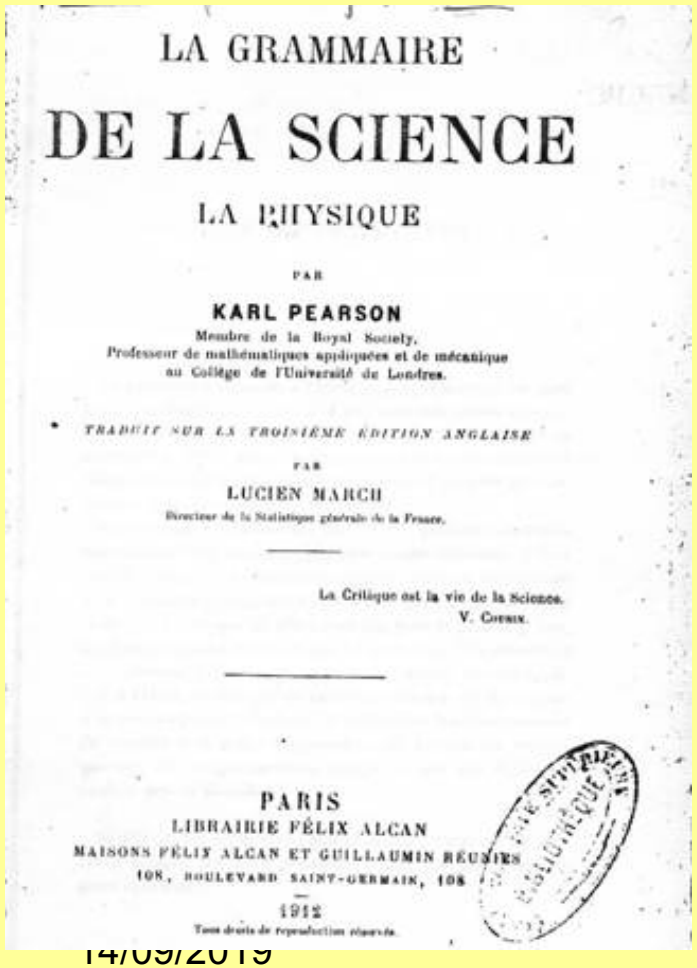
-1892 : premières conférences de statistique, la physique de Mach, la Grammaire de la Science

-1895 Fondation du Biometric Laboratory et publication des Contributions mathématiques à la théorie de l'Evolution" (1894-1912)

- 1901 Fondation de Biometrika

- 1912 Chaire d'Eugénique + Département de Statistique

Karl Pearson 1892



LA GRAMMAIRE

DE LA SCIENCE

LA PHYSIQUE

PAR

KARL PEARSON

Membre de la Royal Society,
Professeur de mathématiques appliquées et de mécanique
au Collège de l'Université de Londres.

TRADUIT SUR LA TROISIÈME ÉDITION ANGLAISE

PAR

LUCIEN MARCH

Directeur de la Statistique générale de la France.

La Critique est la vie de la Science.
V. Cousin.

PARIS
LIBRAIRIE FÉLIX ALCAN
MAISONS FÉLIX ALCAN ET GUILLAUMIN RÉUNIES
108, BOULEVARD SAINT-GERMAIN, 108
1912

Tous droits de reproduction réservés.

-La science ne traite que des phénomènes, et se montre "*agnostique quant au suprasensible*"

-"*Pour nous, le monde réel réside dans des constructions mentales issues de nos perceptions sensibles que nous projetons hors de nous même et que nous appelons des phénomènes, et non dans les obscures "choses en soi"*

-*Une loi, au sens scientifique, ne fait que décrire en sténographie mentale, la suite de nos perceptions.*

-Puisque "*rien dans l'univers ne s'est jamais répété et ne se répétera jamais exactement"* l'expérience ne révèle, derrière l'infini des individualités, que des "groupes de choses semblables".La causalité n'est qu'une fiction conceptuelle. Ce qui est premier c'est **la contingence, la variabilité, la corrélation**.

Usages du coefficient de corrélation par Pearson

Nature et Nurture (1910)

TABLE II. STRENGTH OF NATURE.

PARENTAL RESEMBLANCES.

Physical Characters.

Pair.	*Organ.*	*Correlation.*
Father and Son	Stature	·51
,, ,,	Span	·45
,, ,,	Forearm	·42
,, ,,	Eye Colour	·55
Father and Daughter	Stature	·51
,, ,,	Span	·45
,, ,,	Forearm	·42
,, ,,	Eye Colour	·44
Mother and Son	Stature	·49
,, ,,	Span	·46
,, ,,	Forearm	·41
,, ,,	Eye Colour	·48
Mother and Daughter	Stature	·51
,, ,,	Span	·45
,, ,,	Forearm	·42
,, ,,	Eye Colour	·51

Pathological Characters.

Parent and Offspring	Pulmonary Tuberculosis (Pearson)	·40 to ·60
,, ,,	Pulmonary Tuberculosis (Goring)	·43 to ·62
,, ,,	Insanity (Heron)	·53
,, ,,	,, (Goring)	·47
,, ,,	Deaf-mutism (Schuster)	·54
,, ,,	Corneal Refraction (Barrington)	·60

Mental Characters.

Father and Son	Ability (Oxford Class Lists, Schuster)	·49
,, ,,	Intelligence (Family Records, Pearson)	·58
	Mean Parental Correlation	·49

TABLE III. STRENGTH OF NURTURE.

Characters Dealt With.	*Correlation.*
Keenness of vision and home environment as measured by cleanliness of body and clothing	+·07
Eye disease and overcrowding	+·05
,, ,, economic condition of home	+·03
,, ,, physical ,, parents	−·06
,, ,, moral ,, ,,	+·02
Myopia and age at which child begins to read	−·08
Liability to phthisis and destitution	+·02
Keenness of vision and number of persons per room	−·10
Myopia ,, ,, ,, ,,	−·07
Moral state of parents and refraction of offspring	−·09
Physical ,, ,, ,, ,,	·00
Economic condition of home and refraction of offspring	−·05
Moral state of parents and keenness of vision of offspring	−·02
Physical ,, ,, ,, ,,	·00
Economic condition of home and keenness of vision of offspring	−·01
Weight of child and mental capacity	+·04
Stature ,, ,, ,,	+·08
Condition of teeth and mental capacity	+·09
Condition of clothing ,, ,, (Boys)	+·04
,, ,, ,, ,, (Girls)	+·24
State of nutrition ,, ,, (Boys)	+·01
,, ,, ,, ,, (Girls)	+·08
Cleanliness ,, ,, (Boys)	+·14
,, ,, ,, (Girls)	+·07
Glands ,, ,,	+·08
Tonsils ,, ,, (Boys)	−·01
,, ,, ,, (Girls)	+·11
Alcoholism of parent and weight of child	+·06
,, ,, stature ,,	+·06
,, ,, health ,,	−·05
,, father and intelligence of child	−·06
,, Mother ,, ,,	−·04
,, parent and myopia in child	−·12
,, father and eye disease in child	−·08
,, mother ,, ,,	+·06
Acuity of vision and time out of doors	·00 ?
Shortsight ,, ,,	·00 ?
'Unhealthy' trade of father and weight of child	+·04
,, ,, ,, height ,,	+·07
Employment of mother and weight of son	+·11
,, ,, ,, daughter	+·07
,, ,, stature of son	+·14
,, ,, ,, daughter	+·11
,, ,, intelligence of son	−·16
,, ,, ,, daughter	+·12
,, ,, health of child	+·08
Wages of father and weight of child	+·10
,, ,, stature ,,	+·09
Number of rooms and weight of child	+·11
,, ,, stature ,,	+·11
Mean nurture value,	**+ ·03**

A negative sign indicates that an unfavourable condition or environment appears on the basis of the data available to indicate an improvement in the character.

Les définitions opératoires de Pearson

- *"variation. Si une courbe peut être construite, dont l'ordonné y est telle que ydx mesure l'occurrence d'un organe ayant sa taille entre x et x+x dans une population considérable (500 à 1000 ou plus), les constantes qui déterminent la forme de cette courbe pour un organe particulier d'un animal particulier sont appelées constantes de variation, ou plus brièvement, la variation de l'organe en question."*
- *"Corrélation. Deux organes d'un même individu, ou d'un couple d'individus parents, sont dits corrélés si une série du premier organe étant sélectionnée, la moyenne des tailles des seconds organes associés apparaît comme une fonction de la taille du premier organe sélectionné"*
- *"Sélection naturelle. La sélection naturelle séculaire est mesurée par le seul changement dû à la mortalité dans la moyenne et l'écart-type de la courbe de variation quand nous passons d'une génération adulte à la suivante (...) La sélection naturelle périodique est mesurée par les changements dus à la seule mortalité dans la moyenne et l'écart-type de la courbe de variation aux étapes successives de la même génération, après élimination des changements dus à la croissance."*
- *"Hérédité. Etant donné un organe d'un parent et un autre organe (ou le même) de sa progéniture, la mesure mathématique de l'hérédité est la corrélation de ces organes pour les couples parent-enfant."*
- *"Régression. Ce terme marque la part d'anormalité qui revient en moyenne aux descendants de parents d'un degré donné d'anormalité. La mesure mathématique de cette régression de l'espèce est le rapport de la déviation moyenne des descendants de parents sélectionnés par rapport à la moyenne de tous les descendants à la déviation des parents sélectionnés par rapport à la moyenne de tous les parents ."*
- Karl Pearson, 1896, Régression, Heredity and Panmixia, *Phil Trans RS*

3. G.U. Yule 1897 : antithèse de Pearson

Yule introduit pour la notion de **coefficient de corrélation partielle** entre 2 variables corrigées des influences directes des autres variables = corrélation entre les résidus des régressions de chacune sur toutes les autres.

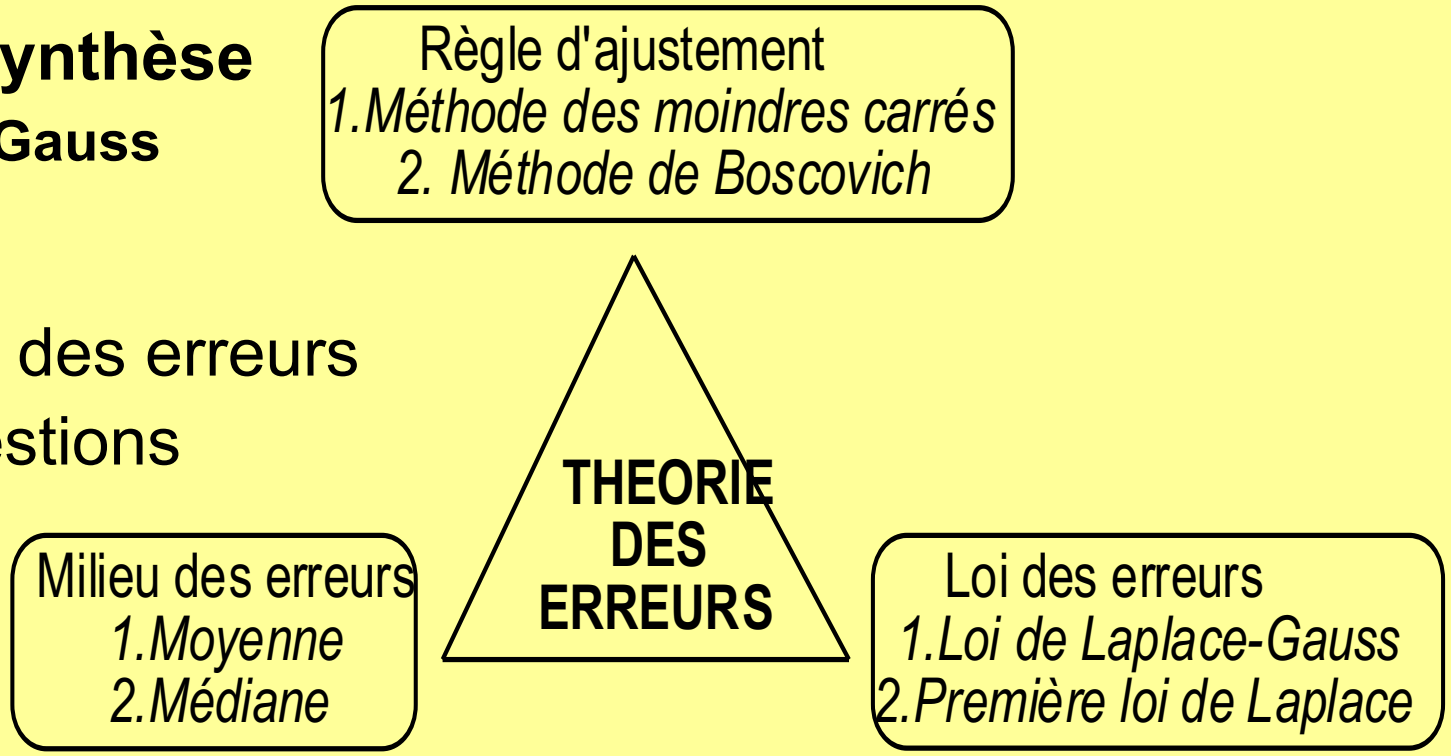
Le coefficient de corrélation est de nouveau (comme chez Galton) un output de la recherche d'une relation de régression linéaire qu'il appelle "causalité statistique".

1°) *Pour qui est familier avec la théorie des erreurs, il sera évident que la méthode n'est que l'application aux objets de recherche statistique de la méthode bien connue des moindres carrés. Il est impossible par conséquent de séparer entièrement la littérature spéciale à la théorie de la corrélation de celle qui touche la théorie des erreurs et la méthode des moindres carrés[1].*

2°) *Comme la forme de la distribution des fréquences donnée par la loi des erreurs n'est pas commune dans les statistiques économiques, il est important d'obtenir la formule de corrélation et ses propriétés sans avoir recours à la distribution de fréquence.*

Yule démontre en 1897 que les droites de régression définies comme lieu des moyennes liées peuvent se déduire directement du principe des moindres carrés puisque la droite des moindres carrés coïncide avec la droite de régression si celle-ci est linéaire (ce qui inclut le cas normal), et s'en approche le plus possible si elle ne l'est pas. Cette nouvelle approche reçoit l'appui d'Edgeworth (1902 et 1908) et de Bowley (dans son traité), il ne reste plus qu'à montrer sa pertinence et son efficacité en économie. (Paris 1909)

4. Théorie des erreurs : La synthèse de Laplace-Gauss



- La théorie des erreurs posait 3 questions

- Les travaux de Laplace et Gauss valident une solution qui **combine de trois résultats**:
  - La Loi des erreurs est la loi de Laplace-Gauss dite plus tard loi normale
  - Le Milieu à prendre entre plusieurs observations pour estimer une grandeur est la moyenne
  - La méthode de compensation des erreurs pour ajuster une droite est la méthode des moindres carrés.

THEORIE DES ERREURS		BIOMETRIE
Gravitation universelle (Newton)	**Théorie**	Evolution (Darwin)
Navigation, Cartographie	**Enjeux**	Eugénisme/Hygiénisme
La figure de la terre	**Exemple**	Loi de l'hérédité
Boscovich, Legendre, Laplace,Gauss	**Acteurs**	Galton , Pearson
1750-1820	**Période**	1885-1900
Objet unique: la géode	**Référent**	Individus multiples parents
Observations indép. répétées y = longueur d'un degré : aléatoire x = sin² latitude : *déterminée*	**Mesures** **y** **x**	Statistiques sur un échantillon y = taille fils : aléatoire x = taille mid-parent : *aléatoire*
Petit nombre d'observations	**n**	Grand nombre d'observations
Newton ⇒Géode elliptique	**Modèle**	Darwin⇒aucune formulation
$y_i = \underline{\bar{y}}_i + \varepsilon_i = ax_i + b + \varepsilon_i$	**Syntaxe**	$y_i = \underline{\bar{y}}_i + \varepsilon_i = ax_i + b + \varepsilon_i$
Ajustement fonctionnel de y_i La fonction découle du modèle Approximation linéaire de la fonction	$\underline{\bar{y}}_i$ $\underline{\bar{y}}_i = f(x_i)$ **Linéarité**	Espérance conditionnelle de y/x_i *Régression* qui découle de la loi Découverte a posteriori et liée à la normalité
Vraie valeur de paramètres du modèle (aplatissement)	**a,b**	Influences à interpréter a=coefficient de régression (<1)
Précision = Erreur de mesure sur y_i +erreur de spécification sur le modèle	ε_i **(signification)**	Variabilité individuelle intrinsèq. + démultiplication du référent
Inutile pour Boscovich ou Gauss (21) Justifie les M.C. pour Gauss (1809)	**Normalité des ε_i**	Hypothèse fondamentale d'une "surface de corrélation normale"
Faible (fluctuation d'erreur) Une hypothèse simplificatrice Règle d'ajustement : moindres carrés	$\sigma^2(\varepsilon_i)$ Homoscédasticité Son minimum	Forte (variabilité descendance) Une conséquence Pas d'interprétation particulière
Qualité de l'ajustement	**$(1-r^2)$**	Réduction de la variabilité
Pas d'interprétations particulières	**r**	Coefficient de corrélation et Pente de la droite de régression en coordonnées réduites

5. Petit détour par l'économétrie

- La biométrie de Pearson fait référence à la théorie des erreurs pour mieux s'en séparer : Voir le texte de K. Pearson, Note on the history of Correlation, *Biometrika 13*. 1920, et Kendal et Pearson *Studies…*, 1970.
- Voir le tableau précédent et la réinterprétation sémantique des différents éléments du modèle linéaire. C'est le même formalisme qui est interprété tout à fait à rebours en s'appuyant sur l'idéalisme de Pearson. Yule tente de s'en dégager
- A la même époque, les années 1910-1920, on est dans une situation similaire en économétrie. La tentative pour mobiliser les concepts du modèle linéaire de la théorie des erreurs se heurte à de grandes difficultés sémantiques. Les économistes peuvent ils reprendre le modèle de la corrélation / régression des astronomes? Ou celui des biométriciens? Pour ne prendre que la notion de covariation entre séries chronologiques, qui diffère par nature d'une corrélation entre erreurs d'observation astronomique tout autant que d'une corrélation biométrique entre tailles de deux parents, le débat fait rage, comme je le montre par un petit détour vers les travaux de Moore et March sur les baromètres économiques.

H.L. Moore 1914

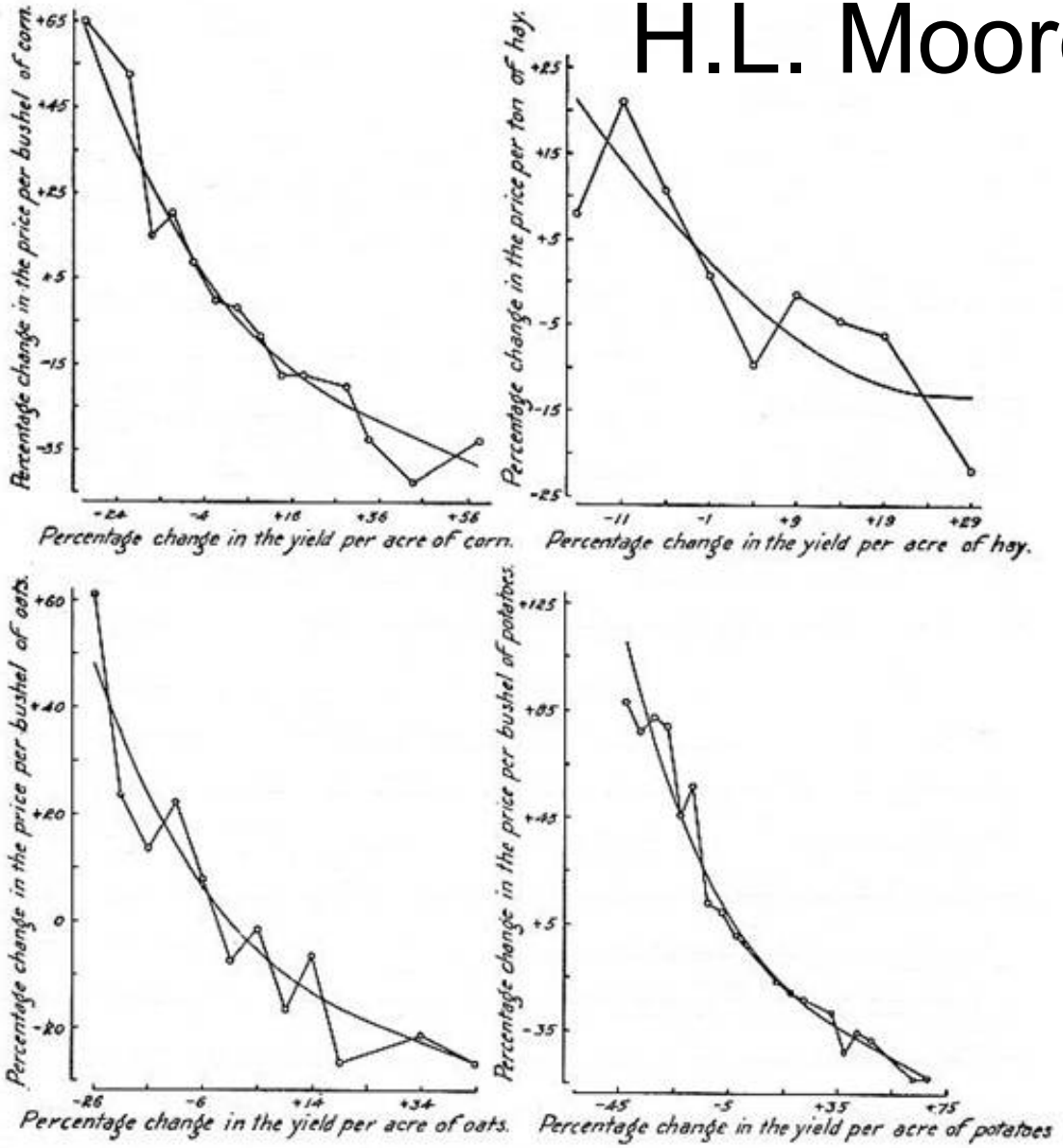


FIGURE 21. The relation between the price and the yield per acre of the several crops.

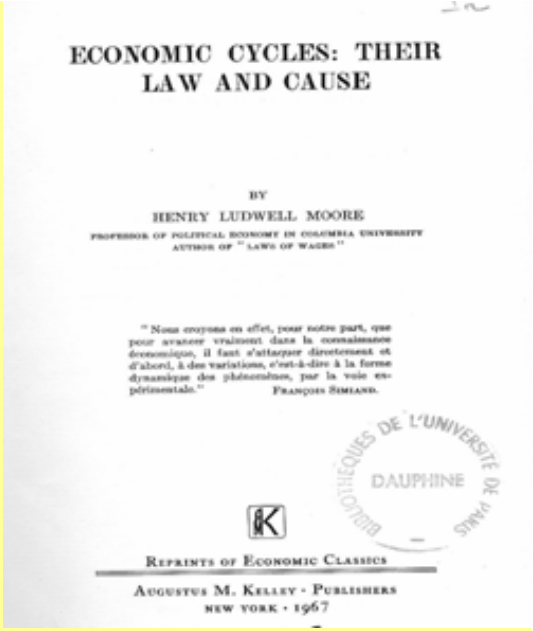
When the origin is at (0, 0), the equations are

For corn, $y = .17 - 1.2989x + .01892x^2 - .000137x^3$.

For hay, $y = 1.17 - 1.0215x + .01549x^2 + .00009x^3$.

For oats, $y = -1.49 - 1.1346x + .02324x^2 - .000238x^3$.

For potatoes, $y = .49 - 1.4863x + .01993x^2 - .000141x^3$.


ECONOMIC CYCLES: THEIR LAW AND CAUSE

BY

HENRY LUDWELL MOORE

PROFESSOR OF POLITICAL ECONOMY IN COLUMBIA UNIVERSITY
AUTHOR OF "LAWS OF WAGES"

"Nous croyons en effet, pour notre part, que pour avancer vraiment dans la connaissance économique, il faut s'attaquer directement et d'abord, à des variations, c'est-à-dire à la forme dynamique des phénomènes, par la voie expérimentale." FRANÇOIS SIMIAND.

REPRINTS OF ECONOMIC CLASSICS

AUGUSTUS M. KELLEY · PUBLISHERS
NEW YORK · 1967

Pour plusieurs produits agricoles, Moore a ajusté une courbe polynomiale sur les valeurs observées des prix et des quantités (en variations relatives). Il obtient ainsi des fonctions de demande décroissantes conformes à la théorie néoclassique.

Une courbe de demande croissante?

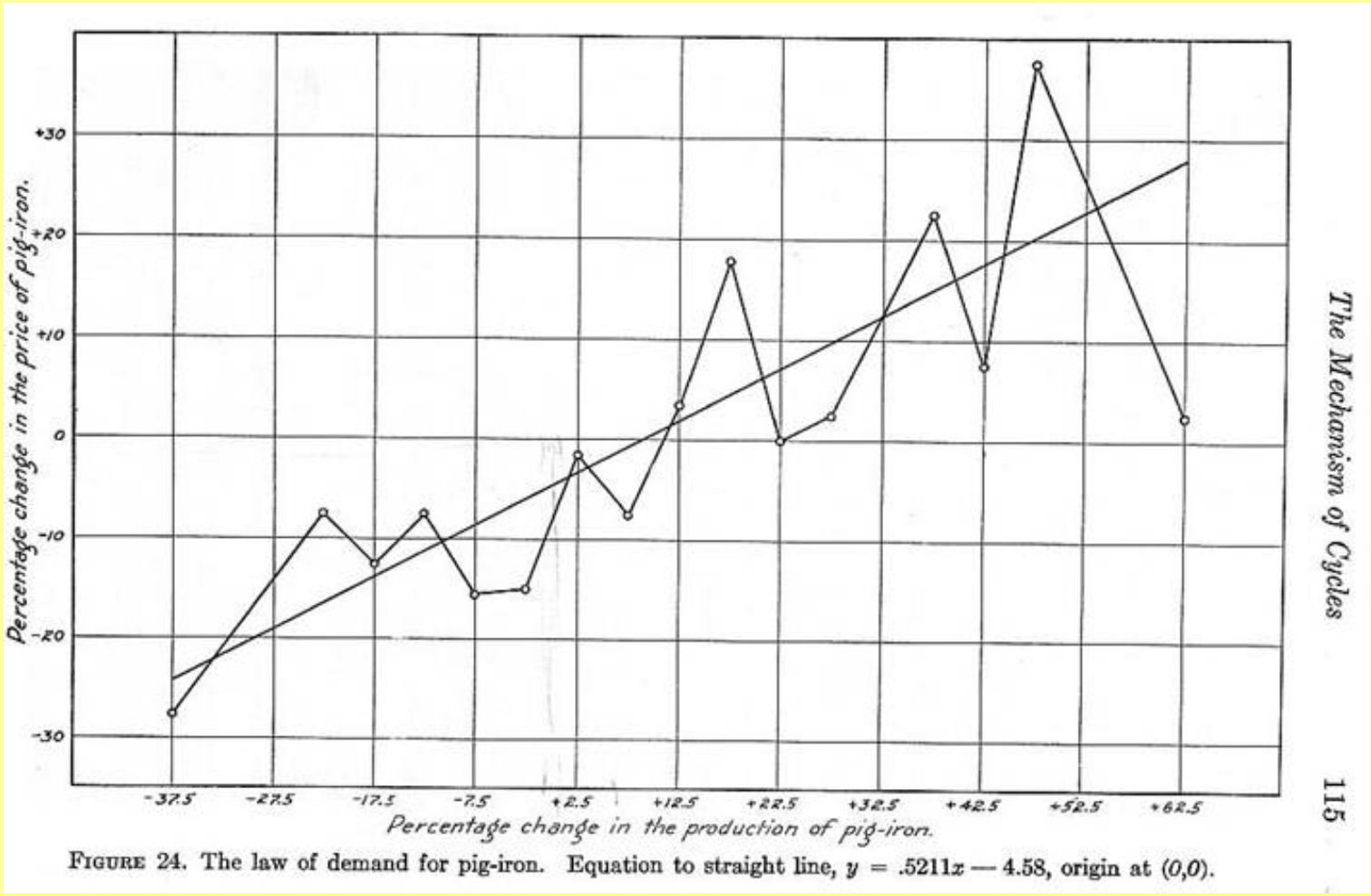


FIGURE 24. The law of demand for pig-iron. Equation to straight line, $y = .5211x - 4.58$, origin at $(0,0)$.

Mais dans le cas d'un bien industriel (la fonte), la même méthode débouche sur une courbe croissante. Ses collègues disent qu'il s'agit d'une courbe d'offre. Moore prétend qu'il s'agit d'une courbe de demande dynamique, différente de la demande statique des théoriciens valable toutes choses égales par ailleurs (*ceteris paribus)*

Le problème de l'identification

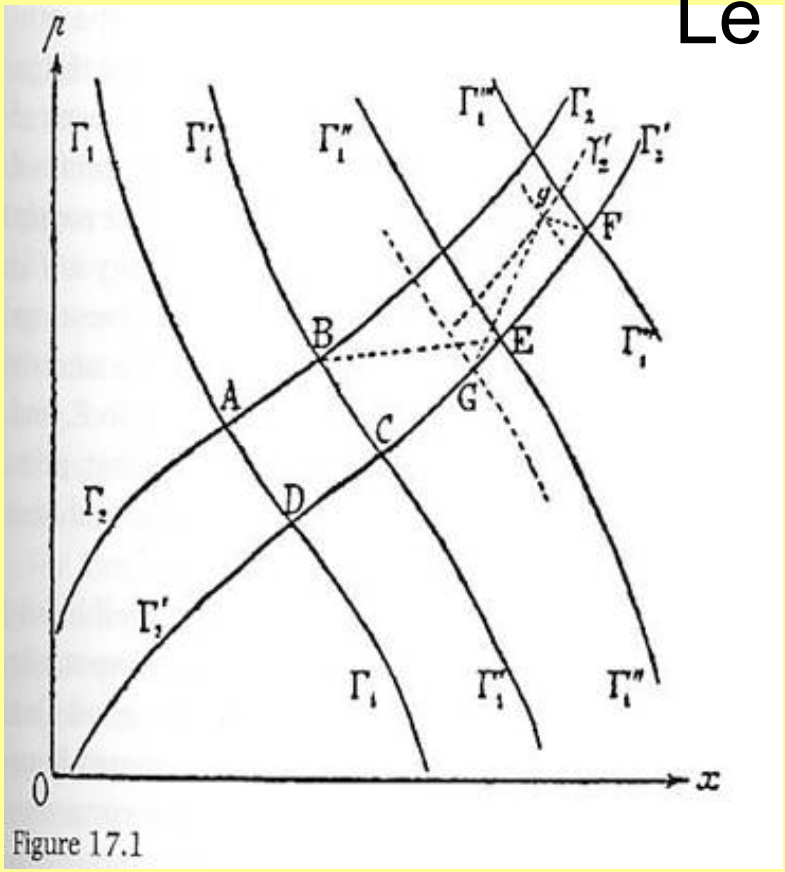
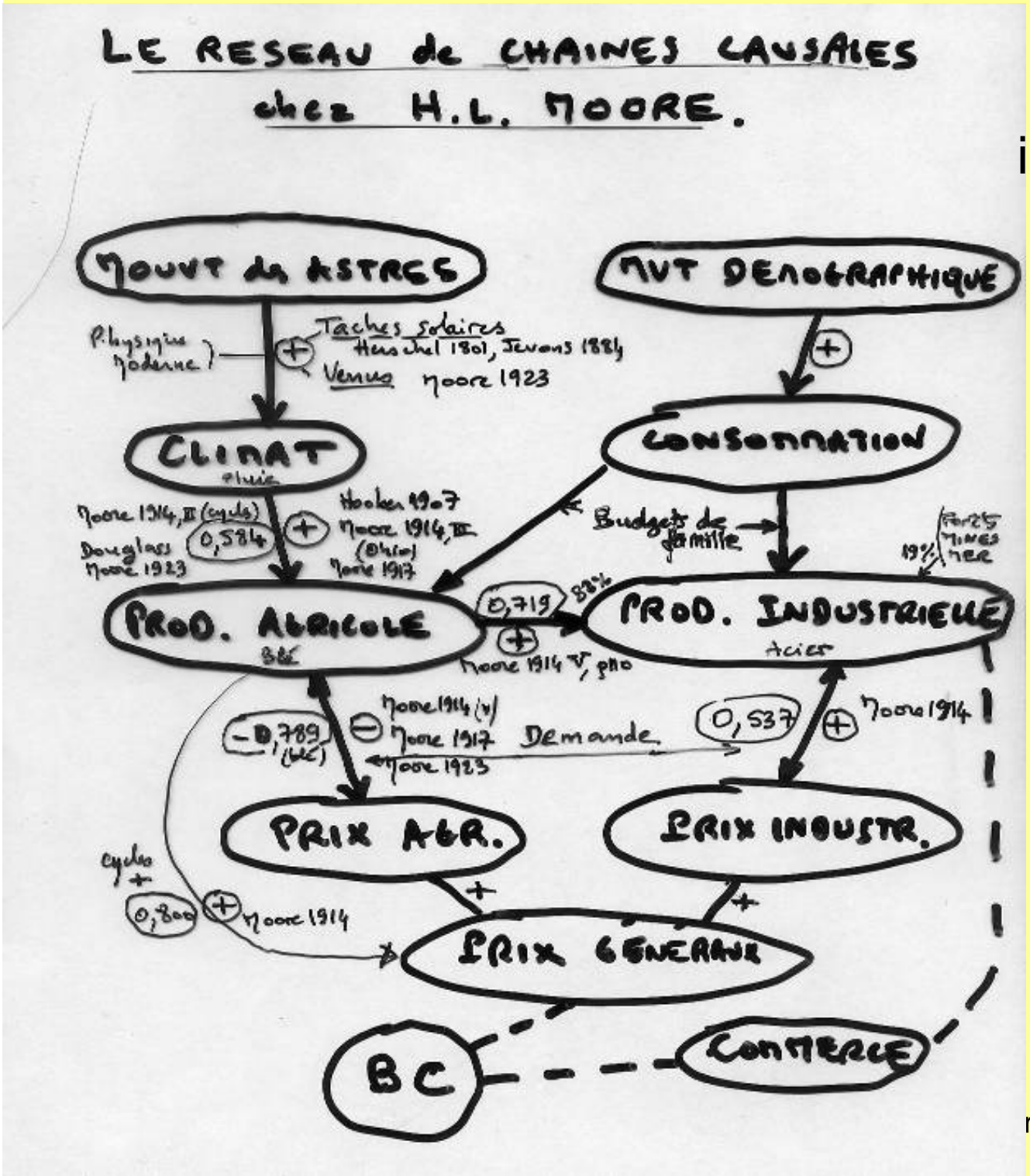


Figure 17.1

Le français Lenoir montre bien dans sa thèse de 1913 que la courbe qui ajuste la suite des points d'intersection d'un ensemble de courbes d'offres (croissantes) et de demande (décroissante) n'est pas en général interprétable en terme de courbe d'offre ou courbe de demande sauf cas limite où l'un des systèmes de courbes est stable et se réduit à une seule courbe.

C'est le problème épineux de l'identification pas toujours possible des relations fondamentales d'une économie à partir des seules observations statistiques des points d'équilibres observés. Le économètres devront résoudre se problème par la notion de modèle à équations simultanées (Moore 1925, Wright 1928, Tinbergen 1930, Haavelmo 1933)

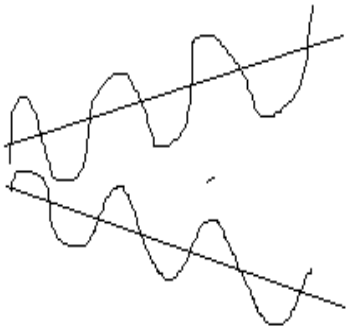


H.L.Moore : interprétation causale des corrélations

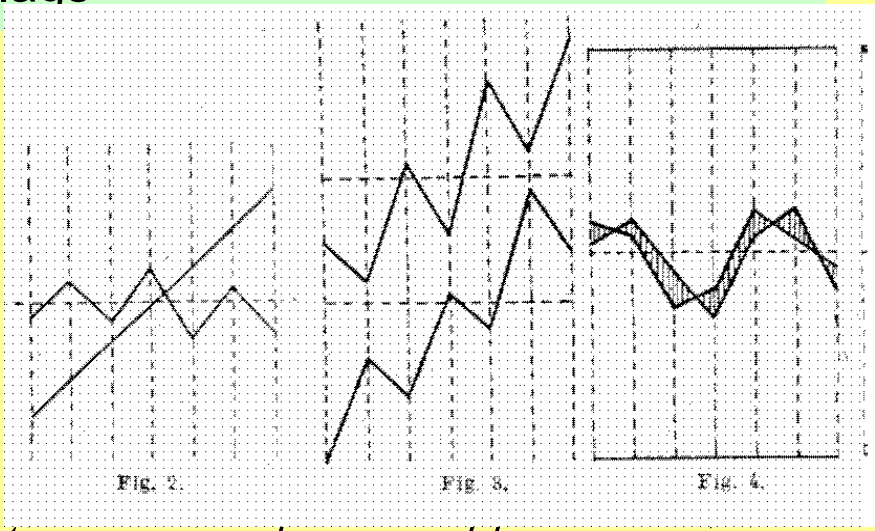
H.L. Moore a développé un schéma causal d'explication des cycles économiques qui remonte au cycle de la planète Venus en passant par les cycles climatiques. Chaque liaison statistique de ce réseau causal a été l'objet d'une évaluation numérique.

Taux de mariage et prix du blé : les pièges de la covariation

- William Farr : Liaison inverse (raisonnement malthusien)
- Willian Ogle (1890) : liaison directe (preuve graphique)
- Hooker (1891) repris par Bowley (1901) r<0 avant 1870 et r>0 après 1870
- Hooker (1901) : nécessité de séparer corrélation entre tendances et entre oscillations cycliques : moyenne mobile et lags



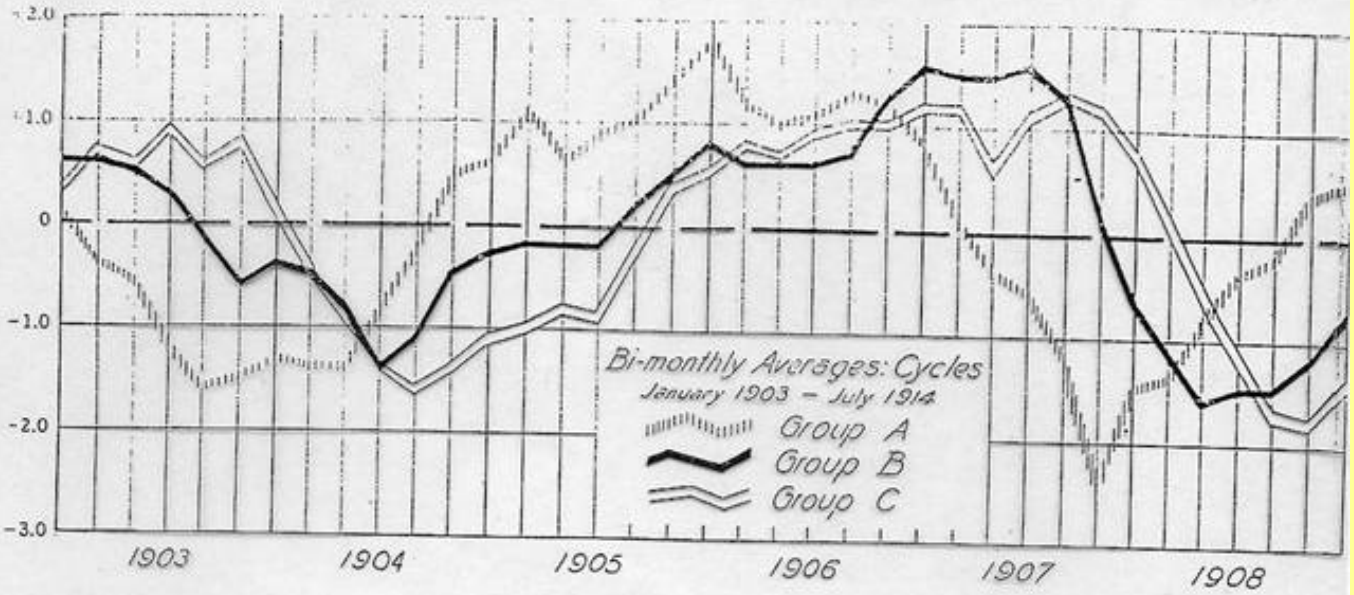
r = 0,18 en valeurs brutes
r = 0,80 en écarts au trend



•Lucien March (1905): : "*Le coefficient de corrélation exprime la ressemblance numerique des variations des grandeurs comparées, sans aucune hypothèse sur le mode de distribution de ces grandeurs*

•March (1928) : covariation tendancielle et covariation différentielle

Le Baromètre de Harvard

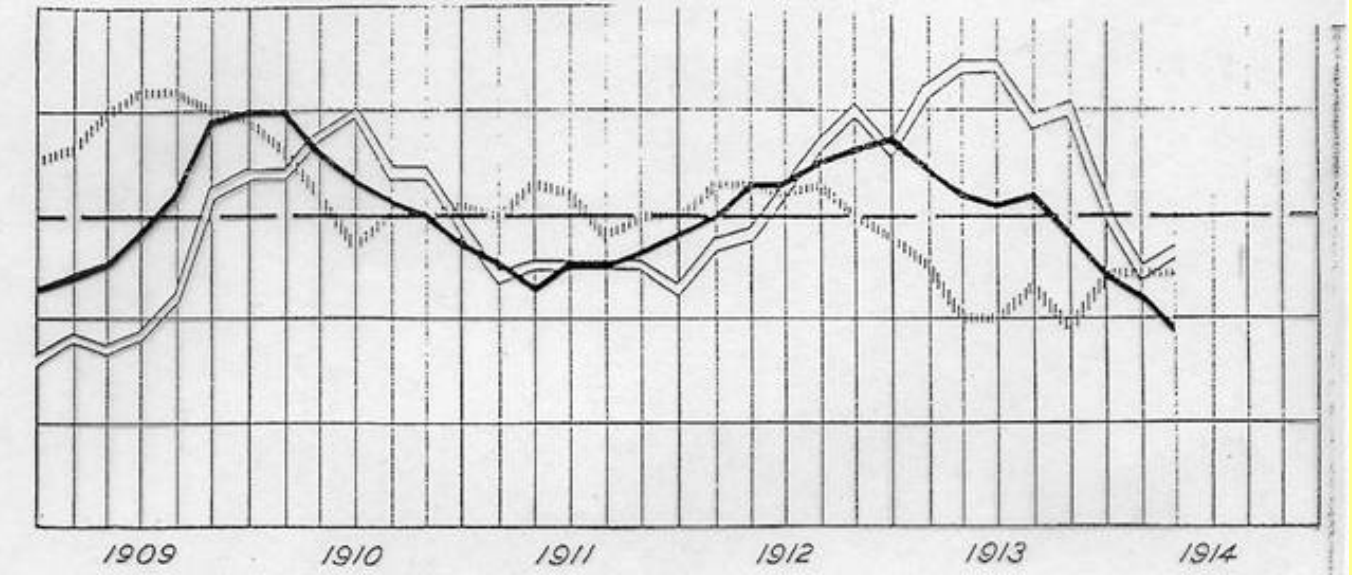


BIMONTHLY AVERAGES OF CYCLES OF GROUPS A, B, AND C

GROUP A. YIELD OF TEN RAILROAD BONDS;* PRICE OF INDUSTRIAL STOCKS; PRICE OF TWENTY RAILROAD STOCKS; NEW YORK CLEARINGS.

GROUP B. PIG-IRON PRODUCTION; OUTSIDE CLEARINGS; BRADSTREET'S PRICES; BUREAU OF LABOR PRICES; RESERVES OF NEW YORK BANKS.*

GROUP C. RATE ON FOUR-TO-SIX MONTHS PAPER; RATE ON SIXTY-TO-NINETY DAY PAPER; LOANS OF NEW YORK BANKS;* DEPOSITS OF NEW YORK BANKS.*

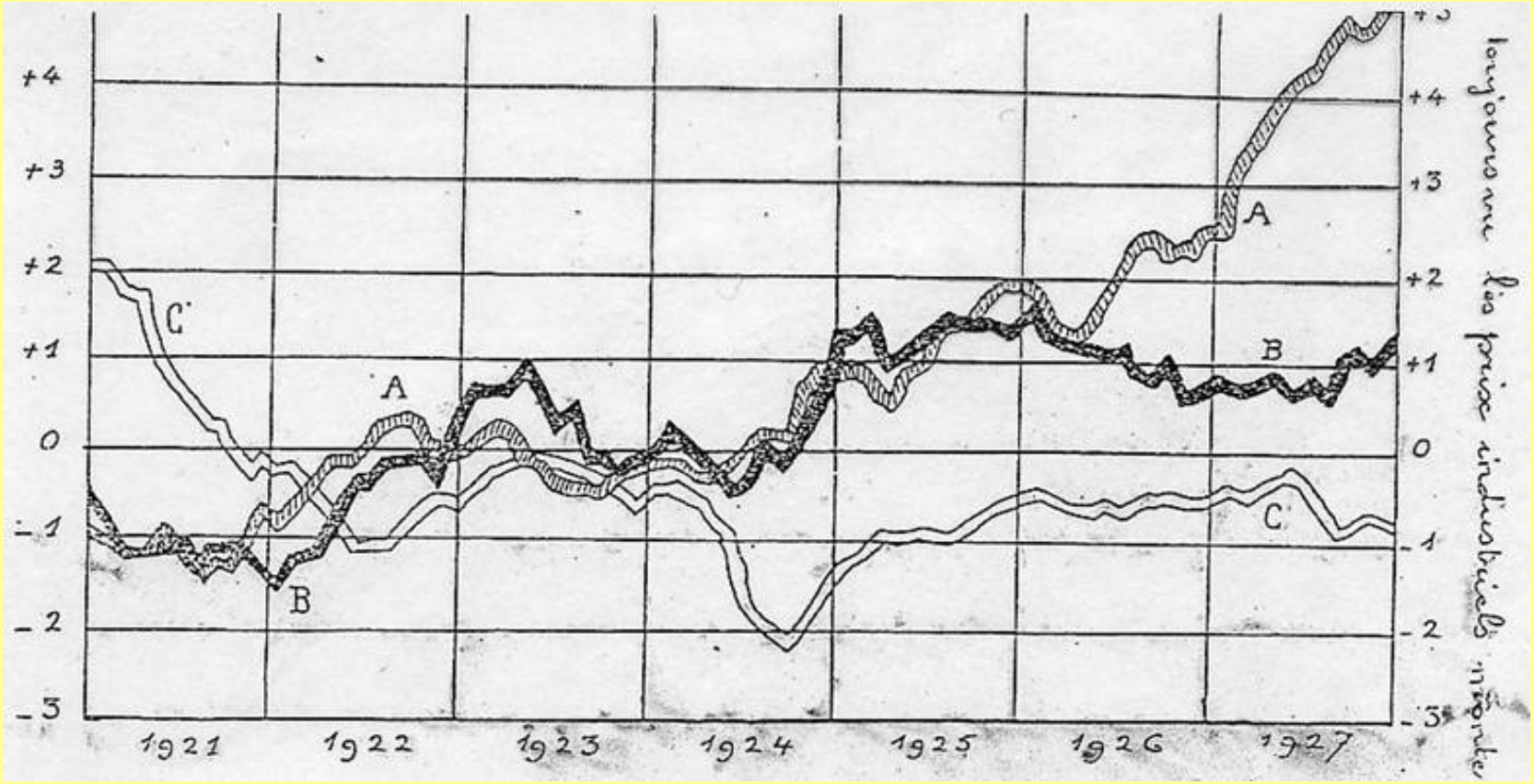


Le Baromètre de Harvard a été conçu dans cette Université par Warren Persons et publié, avec toute sa méthodologie à base de corrélation, dans la revue créée par lui : *Review of Economic Statistics*

Il comprend 3 groupes de séries : A représente l'évolution du marché des valeurs boursières (Speculation), B celle du marché des biens et services (Business) , C celle du marché de la monnaie (Money). Le premier joue le rôle d'indicateur avancé permettant la prévision.

Ce baromètre a servi de modèle à beaucoup d'autres aux Etats-Unis et dans toute l'Europe. Un grand nombre d'instituts de conjoncture se sont créés dans les années 1920 qui avient pour principal produit un baromètre de ce type.

Baromètre de Harvard (2)



Mais dès 1925 les 3 courbes divergent et des économistes se mettent à critiquer le baromètre de Harvard. En 1929 il s'avèrera qu'il fut incapable d'anoncer la formidable crise économique qui allait succéder à la crise boursière et provoquer la faillite de milliers d'entreprises et le chômage de millions de travailleurs.

L'analyse minutieuse de cet échec de la méthode des corrélations et des baromètres dans le domaine de la conjoncture a été la première tâche des économistes après 1930. Ce fut la base d'une nouvelle approche : l'économétrie

La méthode de la corrélation déstabilisée

- En 1921 Yule avait montré que la méthode des différences avait pour effet fâcheux de filtrer certaines périodes (2)
- En 1926 Yule dénonce les "spurrious correlations" sur données en coupe puis montre que la corrélation de deux séries chronologiques peut être totalement artéfactuelle :
  - En aucun cas les observations successives ne peuvent être considérées comme une suite de tirages indépendants de même loi.
  - L'autocorrélation sérielle de chaque série peut produire artificiellement une corrélation élevée entre elles.
- En 1927 Yule s'attaque à l'analyse harmonique dont il montre qu'elle peut provenir de certaines perturbations. La même année, Slutzky avait mis en évidence les autocorrélations artificiellement provoquées par l'opérateur moyenne mobile.

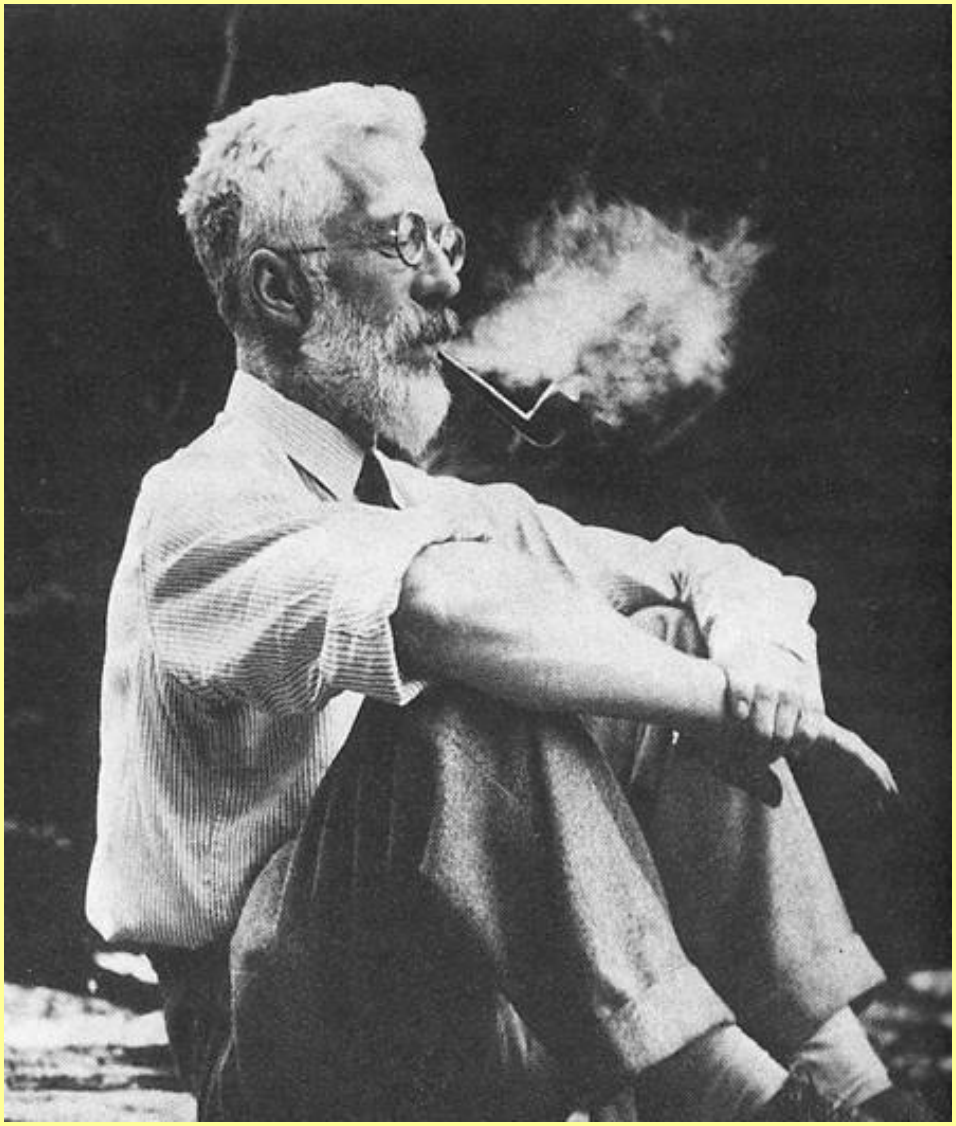
En 1935, le mathématicien Fréchet mène à l'IIS une campagne vigoureuse contre les usages abusifs du coefficient de corrélation pour repérer à la fois l'intensité et la linéarité d'une liaison statistique, et dont la maximisation produit le lag optimal dans les baromètres.

6 Ronald Fisher

Plate 8. March 1929, Fellow of the Royal Society.

Plate 9. Mrs. Fisher with Elizabeth, 1931.

Ronald Fisher



- Vision *probabiliste* : étude avec Student de la distribution du coefficient de corrélation

-- vision *géométrique* de la corrélation : c'est le cosinus de l'angle (x,y)

--- vision *inférentielle* : estimation de la corrélation

--- Ce qui importe c'est la *significativité* de la mesure de corrélation, c'est à dire le rapport à l'écart-type de sa fluctuation.

-- La variance et sa décom-position comme mesure de la variabilité et de la causalité

Biographie Ronald Fisher (1890-1962)

- Ronald dit Ronnie est le benjamin des 7 enfants (dont 1 frère jumeau mort-né) d'un commissaire priseur londonien (Georges), lui-même fils aîné de John et d'une fratrie de 10 enfants.
- Précocement doué, mais mal voyant et insensible « unaware of of the effects of his own behavior» au monde extérieur (suite décès de sa mère en 1904?) ; « he had a deconcerting habit of producing the correct answer without showing you how he had arrived at it »(Box) : maths sans crayon
- 1909-1913 Caïus college à Cambridge. Il fonde une sorte de cercle des poètes disparus (The Wee Frees) et une société d'eugénique at Cambridge (1910) sur le modèle londonien (Galton 1907, président Leonard Darwin)
- 1911 Fisher's talk at 2d meeting of Cambridge eugenic society, édité par Norton et ES Pearson (176) repris à Londres (CP3) : Décadence of civilized races, dénatalité différentielle des élites.
- 1913-1919 Il est 6 ans sans job satisfaisant sauf une année de stage agricole dans une ferme canadienne, une année à la Mercantile and Investissment Cy à la City et 4 ans comme prof. de math et physique à Rugby puis Bradville college in Kent: « Fisher was a poor teacher » (Source Joan Fisher-Box 1978 *The life of a scientist*, et Yates et Mather, *Collected Papers* 1971-74.

Biographie de Ronald Fisher (suite)

- En 1915 il se présente à la circonscription mais est refusé pour cause de malvoyance. Très affecté, il ne cessera de tenter sa chance jusqu'en 1918.
- Avril 1917 Mariage avec Eileen Guiness, sœur de son amie Gudruna, veuve avec 2 enfants. Ils s'installent à Great House Cottage où Fisher se passionne pour le farming (poultry, garden, pig) mais en réalité il est accaparé par ses enseignements et ce sont les deux femmes qui gèrent la ferme. Vie austère et solitude : « no outing, no neighboors, no radio, no journals » dit Cox.
- Ronald et Eileen auront 8 enfants dont 6 filles. Katie, morte accidetellement en 1920. Son frère Alwin est mort en France en 1915 comme soldat. Une sœur meurt en 1917 et son père en 1920. Gudruna les quitte en 1920.
- A la recherche d'un job en 1919, il essuie plusieurs échecs au Caire, en Nouvelle Zélande, à Cambridge avant d'obtenir, via le botaniste Horace Brown, un emploi précaire à Rothamsted … qu'il gardera jusqu'en 1933. Il refusera l'offre de travail de K. Pearson au *Galton Laboratory* assortie de conditions « castratives » inacceptables : teach and publish only what Pearson approved.
- En 1933, à la retraite de KP, Fisher occupera la chaire d'Eugénique de London College (conflits avec Egon Pearson). Retour à Rothamsted pendant la guerre. Balfour professor of genetics at Cambridge in 1943-1957. Retraite à Adelaïde

7. R. Fisher son œuvre statistique (extraits) 1935

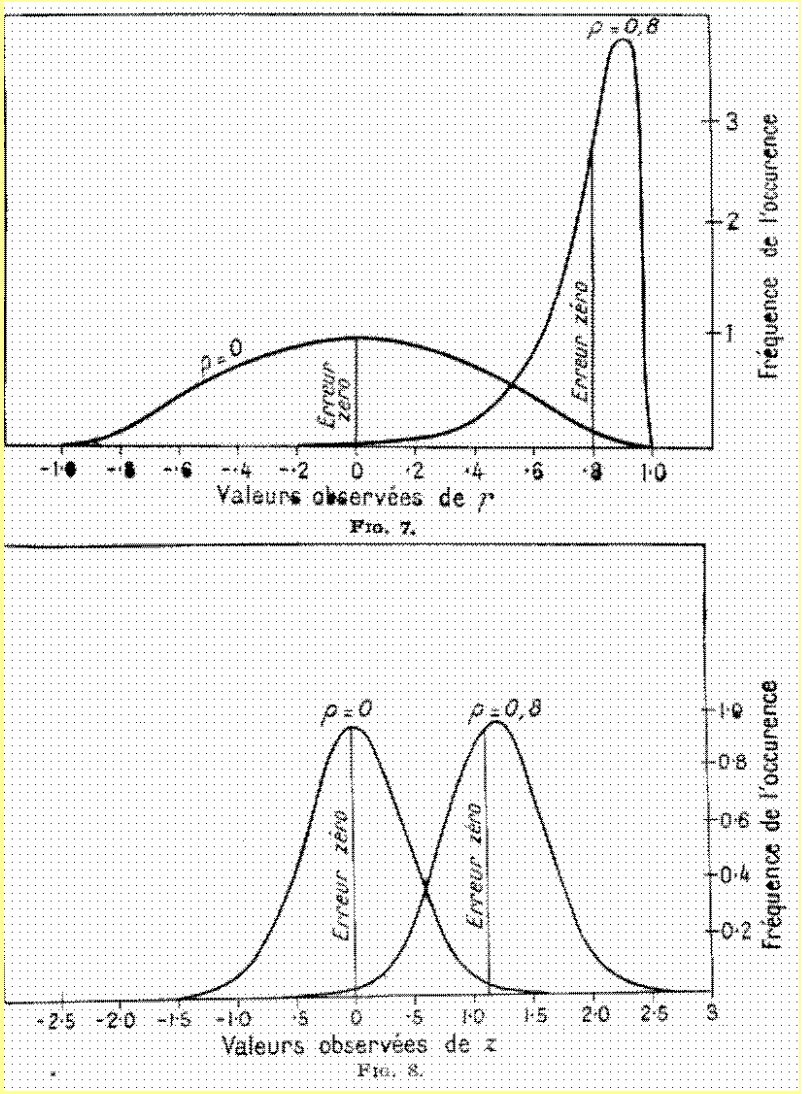
- 1912 CP1 On an absolute criterium for fitting frequency curves
- 1913 CP2 Application of vector analysis to geometry, *Messeng Math*
- 1914 CP3 Some hopes of a eugenist, *Eugen. Review*
- 1915 CP4 Frequency distribution of the correlation coeff., *Biomretrika*
- 1915 CP6 The evolution of sexual preference*, Eugen. Review*
- 1917 CP8 Positive Eugenics, *Eugen. Review*
- 1918 CP9 **The correlation between relatives on supposition of Mendelian inheritance**, *Trans. Roy. Soc. Edinb.*
- 1918 CP10 The causes of human variability, *Eugen. Review*
- 1920 CP12 Mathematical examination of the method of determining the accuracy of an observation by the mean error and the mean square error
- 1921 CP14 On the probable error of a coeff. de correlation, *Metron*
- 1921 CP15 Studies in crop variation I , *J.Agri.Sci (*1923 CP32 II)
- 1922 CP18 On the mathematical foundation of theoretical statistics, JRSS
- 1922 CP20 The goodness of fit of regression formulae JRSS
- 1925 *Statistical Methods for Research Workers, Edinb, Oliver & Boyd,14 ed*
- 1935 *The Design of Experiments, Oliver & Boyd, 8 ed*

Statistique mathématique et logique inductive

- Armatte, La construction des notions d'estimation et de vraisemblance chez Ronald Fisher, 1988.
- CP1 1912 Méthode du maximum de vraisemblance. Fisher à la recherche d'une méthode statistique inductive démolit pour la 1ere fois (repris en 1937) la méthode de Pearson d'ajustement des moments théoriques et empiriques. (De tels estimateurs sont sans biais mais pas de variance minimale) La maximisation de la vraisemblance d'un échantillon observé est le critère défendu par Fisher . La vraisemblance est malheureusement assimilée en 1912 (Par KP et d'autres) à la probabilité d'un échantillon et la méthode à celle de la probabilité inverse déduite des travaux de Bayes.
- De 1922 à 1930 il ne cessera de rejeter cette méthode (qui repose sur des formules bayesiennes exactes mais sur le postulat de répartition uniforme de la loi a priori (principe de raison insuffisante) et sur la maximation du mode de cette distribution, un critère arbitraire . Il veut que l'on distingue vraisemblance inductive (non intégrable) et probabilité, et que l'on cesse de probabiliser l'ensemble des valeurs du paramètre.
- La justification viendra ensuite avec sa théorie de l'estimation (1922): un tel estimateur est central dans la stratégie de résumé de l'information, car il est à la fois convergent (consistent = asympt. sans biais) , efficace (efficiency = variance mini) et suffisant ou exhaustif = qui conserve une certaine information portée par l'échantillon.
- CP124 1935 Logic of inference «L'étude du raisonnement inductif est l'étude de l'embryo-logie de la connaissance ». (en conflit avec ES Pearson, Bowley, Jeffreys, Neyman…)

La distribution exacte de r

- CP4 1915: Distribution exacte du coeff de corrélation sur les pas de Bowley (1906) et W.Gosset (alias) Student (1908). Pearson avait supposé que la statistique r était distribuée asymptotiquement selon une loi normale donc symétrique d'espérance ρ et d'écart-type $(1- \rho^2)\sqrt{n}$.
- Fisher établit géométriquement (CP2, CP4) que la loi de r est asymétrique d'espérance bien inférieure à r. Il propose alors 2 changements de variable $t=r/\sqrt{(1-r^2)}$ suit la loi de Student, et dans un second papier, z = $than^{-1}(r)$ distribuée quasi normalement.
- Pearson accepte de publier le premier papier dans Biometrika mais refuse de publier le second: « dans les conditions éditoriales et financières actuelles, je suis au regret d'exclure tout ce que je considère de mon propre jugement, comme erroné car je ne peux accepter la controverse » et publie en 1917 sa propre table de r et une critique de la méthode du maximum de vraisemblance. Ce second papier de Fisher sera publié dans Métron en 1921 (CP14)



Corrélation et analyse de variance

- Comme on l'a montré plus haut chez Galton, r^2 est dans la régression de y en x le rapport entre la variance des moyennes liées de y et la variance totale de y et $(1-r^2)$ le rapport des variances intraclasse à cette même variance totale.
- CP9 1918 *Correlation between relatives.* L'hypothèse de départ du papier est qu'il convient d'analyser les causes de variabilité par le biais de la variance car les variances de causes multiples indépendantes sont additives, ce qui nous permet d'assigner à chaque cause une fraction de la variance totale, **tout en évitant « the loose phrases about the percentage of causation** » dit Fisher. Il ne cite aucune source.
- CP15 1921 *Studies in crop variation envisage la* décomposition de la variance totale de taille entre n familles de k frères en variance inter-classes et variance intra-classes (due à la famille, cause commune) . Repris Fisher 1925 : il écrit le modèle $x_{ij}=u_i+v_{ij}$ u_i est l'effet familial et $r = V(u_i)/V(x_{ij})$ Randomization, Essais cliniques randomisés, eval polit pub JPAL
- CP18 1922 (On the math foundation of Stat) fournit les bases théoriques des statistiques mathématiques avec 4 concepts : réduction des données, information de Fisher, population hypothétique infinie = famille de lois paramétriques, induction optimale sous forme d'estimation et de test d'hypothèse, résumés, convergents, efficaces, exhaustifs.
- CP20 1921 explore les distributions de probabilité des résumés d'une régression linéaire : coefficients (Student), variances totale expliquée et résiduelle (Chi^2)
- Attention la part de variance de y expliquée globalement par 3 variables x1 x2 x3 n'est pas décomposable sur chacune, et elle dépend des autres variables explicatives du modèle. Voilà qui relativise la notion d'héritabilité. Pas de sens absolu

Biométrie et génétique

- Mendel (1865) redécouvert en 1900 par Tschermak, Correns et de Vries. La notion de gène par Johansen 1911 et l'association de gènes au sein de chromosomes (linkage) par Bateson 1908, Punnett, et Morgan 1911.
- Box : à l'époque, il n'y a pas de connexion logique entre darwinisme et génétique : les caractères, y compris acquis; sont hérités, la variation continue se combine avec une sélection naturelle indépendante (hasard) pour opérer une évolution progressive. Les inventeurs de la génétique mendélienne veulent renouveler ce credo, mais différemment : Bateson penche pour une hérédité et une évolution discontinues. De Vries pense que la mutation joue un rôle essentiel, plus important que la sélection.
- Pearson 1903 insiste sur la sélection qui agit continument par petites inflexions. En cohérence avec son idéalisme de la *Grammaire de la science* et sa conviction que les lois (des résumés sténographiques de nos perceptions), sont continues (≠Yule)
- Fisher découvre la génétique via l'eugénisme (idealized) à Cambridge (1909 - 1913) où Bateson enseigne la biologie puis la génétique (1912). *Correlation between relatives* est le premier et l'avant dernier papier de Fisher sur ce sujet. Fisher pense qu'une hérédité discontinue n'est pas contradictoire avec une variation continue. La synthèse entre eugénique, statistique et génétique doit permettre de « révéler la structure logique qui contrôle les faits empiriques. »

8. Fisher 1918 : un programme argumentaire

- On dispose du texte CP9, de la vulgarisation CP10, et de la revue de Moran et Smith 1966 (explicitation, discussion technique)
- Introduction. L'objectf est d'assigner aux différentes causes les % de variance de la population qu'elles produisent. Et cela grâce aux outils statistiques de la théorie des erreurs. Mais cela n'autorise aucune interprétation individuelle (« loose phrases about percentage of causation »)
- Père fils : r = 0,5 et la part de variance de la taille des enfants expliquée par la taille du père (=inter classe ou des moyennes liées) est r^2 = 0;25. Idem pour mère enfant, mais si on prend en compte la corrélation maritale (28%, 17%) la part de variance n'est pas 50% mais 40% (33% pour Pearson 1903). La prise en compte de tous les ancêtres, déduite de la corrélation entre frères (formule en note) ne permet pas de dépasser 54% de la variance.
- Or le résidu ne peut être expliqué par l'environnement r < à 0,1 (Pearson Nature and Nurture) ni par une dominance parfaite, mais par la ségrégation d'un grand nbre de facteurs mendéliens. En supposant ces facteurs indépendants et d'importance égale, et des phases récessives et dominantes d'égale fréquence, Yule (1906) a déjà montré l'effet conjoint de l'environnement et de la dominance dans la réduction des corrélations entre « relatives », sans pouvoir les séparer. Fisher se propose de le faire , sous des hypothèses de plus en plus générales. Conclusion CP10 : Les causes génétiques représentent plus de 95%

Fisher 1918 : un découpage argumentaire (1)

- 1. Cas d'un facteur simple à 2 allèles. Donne 3 phases (zygotes) A1A1, A1A2, A2A2 de fréquences P,2Q,R ayant un effet phénotypique a,d,-a sur x. (loi Hardi-Weinberg) Sous l'hypothèse d'un grand nombre de facteurs, la déviation x est la somme des déviations causées par chaque facteur: (TCL=> normalité) et $\sigma^2=\sum\alpha^2$
- 2,3 Il suppose qu'un seul des facteurs ne donne que des hétérozygotes et il en déduit les modifications des fréquences qui en résultent et la régression due à ce facteur.
- 4,5. Effet de la dominance sur cette déviation. Epistasie = imperfectly additive genetic factors. β^2 est la contribution de la part de variance additive et δ^2 de sa part résiduelle, le total α^2 étant la variance génotypique totale. (il cite Malecot) . Le résultat essentiel (parental correlation for a static population mating in random is simply $r^2/2\sigma^2$) est interprété en terme de réduction des corrélations due à la dominance.
- 6,7. Ce résultat doit être étendu aux parentes collatérales car « déviations from linearity are now themselves correlated». Sauf dans le cas des siblings (enfants de mêmes parents). Pour chaque couple frère-frère, par ex, il recalcule les tables croisées des génotypes et le résultat en terme de corrélation. (Moran matrices) « The brother-brother correlation is therefore exactly intermediate betwween parents-offsprings correlation with and without the same degree of dominance »

Fisher 1918 : un découpage argumentaire (2)

- 8. Cas d'une combinaison de deux facteurs mendéliens, dans **deux locus A et B** non liés, ce qui donne une table de 3x3 = 9 termes de déviation (erreurs) à 4 degrés de liberté. Et la table parents-enfants a 81 termes est réduite à une table 4x4. Moran et Smith en proposent une écriture matricielle
- 9-10-11. Fisher abandonne l'hypothèse du random mating pour **l'assortative mating.** « there will be association between similar phases of differents factors, so that they cannot be treated separately » (et l'équiibre de Hardy-Weinberg ne joue plus). Il introduit un paramètre nouveau μ dans la covariance de la densité conjointe du caractère, et corrige les formules de P,Q,R les fréquences des 3 phases d'un simple facteur.
- 12-13 s'attaque aux 45 types de mariages des 9 types restreints par les 4 conditions homozygotes. Les calculs deviennent fastidieux. Pour obtenir la formule finale du ratio de la variance avec et sans les écarts dus à la dominance
- 14-15. Fisher se propose alors d'étendre ses résultats au cas où chaque facteur contient plus que 2 formes d'Alleles (**multiple allelomorphisme**). On sort dit-il du modèle classique mendelien pour traiter de l'hérédité particulière à Galton, mais dit-il cela ne change rien à la simplicité de nos résultats (6) dans le cas du random mating. Les corrélations parentales et fraternelles gardent les mêmes formules. C'est plus compliqué quand on couple homogamie et multiple allèlom.

Fisher 1918 : un découpage argumentaire (3)

- 16. **Coupling**. Un hétérozygote est le produit de deux paires de gamètes (1.1)x(2.2) ou (1.2)x(2.1) . Il se peut que les gamètes de cet individu ne donnent pas des chances égales à chacune des 4 types mais donnent une préférence aux deux types dont il est issu. Il conclut que le coupling est sans influence sur les propriétés statistiques de la population.
- 17-18. Fisher revient à la question initiale des effets séparés de la dominance et de l'environnement (*arbitrary external causes*). Il utilise la théorie de la régression. Si x= taille observée y = taille pour un standard environnemental et z = effet (indépendant) des facteurs génétiques. Alors Les résultats en terme de corrélation dépendent **de l'interprétation de la corrélation maritale** (sélection consciente ou inconsciente). Cela pose toute la question de l'indépendance des 2 facteurs y et z dans le choix du conjoint.
- **Moran et Smith** : *Fisher tacitly supposes that the effects of environment can be represented by an addition to the measurement which is independent of the genetic value so that there is no 'interaction' between genotype and environment. This environmental deviation' is supposed to be normally distributed with zero mean and constant variance, and is not correlated among relatives*

Découpage argumentaire (4) Ambiguités

- Then, measuring from the mean, we can write
- x = observed value = y (genetic value)+environmental effect
- = z (representative value)+dominance deviation + environmental effect.
- These three terms, the first two of which are sums over the various loci, are mutually uncorrelated. Thus with a large number of loci, the joint distribution of (x, y, z) is trivariate normal, **with z (representative value), y-z (dominance deviation), and x-y (environmental effect) all statistically independent.**
- [cov (x, y) = var(y) = V, cov(x,z) = cov(y,z) = var(z), var(x) = var(y) +var (η),where η is the environmental effect. Then an increase δz in the representative value will on the average increase both the genetic component y, and the observed measurement z, by δz. This is also evident from the above decomposition. Thus we have b_{yx}= var(y)/var(x). $b_{zy} = r^2/(\sigma^2 - A\varepsilon^2)$]
- Now let x, y, z be the values for a father, and X, Y, Z, the corresponding values for his son. Considérons la regression de X, Y, Z, sur x, y, z.
- Son interpretation depend de 3 hypotheses sur la nature de l'assortative mating, selon que l' association est (1) entre les characters observès x; (2) entre les composantes génétiques y; (3) entre les valeurs representatives z.

Découpage argumentaire (4)

- 17,18. Dans le premier cas, on retient pour essentiel comme Pearson, la corrélation μ entre x et X. Il n'y a pas d'autre association que z entre parent et enfant. Et la corrélation parentale est c_1c_2 (1+μ)/2. Dans le second cas, l'association est essentiellement due à y : la corrélation parentale est $(c_1+c_2+Ac_1)/2$. Dans le 3e cas la connexion entre mari et femme porte sur z et est essentiellement due à la fertilité différentielle.
- 19,20,21 Fisher donne in fine des **valeurs numériques** dans le premier cas où la seule base de la connaissance des corrélations maritales et parentales suffit à déterminer les coefficients c_1c_2 et A. Les corrélations entre sib et doubles cousins sont évaluées sous la même hypothèse. Il applique ses formules aux corrélations entre frères publiées par Pearson et Lee. Aux erreurs d'échantillonnage près, « there is any cause of variance in these growth features than genetics differences » Les 46% inexpliqués par les ancêtres se répartissent, pour la taille, en 62% génotypique, 21% par dominance et 17% par l'homogamie
- 22-25 Fisher fait la même chose, très prudemment pour les oncles et cousins. Et il discute de ses résultats sur les effets de la dominance crédités d'une part de variance comprise entre 0,25 et 0,38. (Pearson 0,33)
- 26 Conclusion « The statistical properties of any feature determined by a large number of Mendelian factors has been succefuly elucidated' Grâce à la corrélation fraternelle, il est possible d'établir la dominance et de la distinguer des causes non génétiques comme l'environne-ment. L'hypothèse de facteurs cumulatifs est à retenir. Mais il reste certaines ambiguités sur les causes de la corrélation maritale.

Retour sur la notion d'héritabilité

- L'héritabilité caractérise un phénotype et se mesure statistiquement par la **part de variabilité d'un caractère qui est due à des causes génétiques** (De Vienne et Clerget & Génin). Elle ne se confond pas avec l'hérédité d'un caractère qui suppose un déterminisme individuel génétique. Elle dépend de la population dans laquelle on fait les mesures. Elle est difficilement isolable des facteurs d'environnement et culturels. Elle n'est pas calculable expérimentalement pour les humains.
- **Albert Jacquard (éloge de la différence 1978)** distingue **l'héritabilité des biométriciens** (nos formules de la régression) où elle varie de 0 (pas héritable) à 1 (totalement héritable) de **celle des généticiens** comme Fisher, qui raisonnent sur les génotypes sous hypothèse d'un milieu homogène, en se fondant sur la loi de Hardy-Weinberg (proba ¼, ½, ¼ des 3 génotypes AA, Aa, aa) pour 2 gènes également répartis. L'effet de chaque gène se mesure en terme de **variation ou d'écart** à la moyenne et dépend de la fréquence des gènes dans la population. Jacquart reprend Fisher 1918 (N°17)
- Fisher a proposé d'analyser les écarts entre génotypes en 2 parts : un effet additif des gènes (estimé indirectement par les corrélations entre parents) et un résidu que l'on s'efforce de minimiser (MCO). Si l'on a pu isoler les effets de la différence de génome des effets du milieu (hypothèse lourde d'indépendance), alors l'héritabilité du caractère est égale au rapport entre la variance des effets additifs des gènes impliqués et la variance totale du caractère. Ceci est nommé **héritabilité stricte** par Jacquart. **L'héritabilité au sens large** est donnée par la technique d'analyse de variance qui permet de séparer la part du génotype et la part du milieu en considérant d'abord des individus de même génotype et de milieux différents puis l'inverse. On peut alors écrire V = VG + VM – I (pour les interactions)

Une évaluation de Samuel Karlin

R.A. Fisher And Evolutionary Theory, *Statistical Science* 1992

- ***7.*** *«* **Fisher and polygenic inheritance**. Propose un bon résumé : « In this presentation, the phenotypic variance is decomposed as a sum of independant « additive-genetic » and « dominance variances plus an independant environmental variance. Fisher computed a number of phenotypic correlations of relatives as functions of the variance components and proposed to estimate components of variance from observed correlations and to assess various heritability coefficients *»*
- *Kempthorne O.,1952, An introduction to Genetic Statistics* characterizes the Fisher 1918 paper as « **remarkably difficult to understand so much that it is still under debate** ». But there is **no natural way to define additive genetic variance for a non HW population** in case of assortative mating, especially in minimizing with respect of least squares.
- « To sum up, under assortative mating with a metrical trait, **serious problems arise** from a) the approximations in the treatment of nonadditivity; b)the definitions and the interpretations of additive genetic and dominance variances; c) the lack of a meaningful analysis of variance for non-Hardy-W populations; d) a hierarchy of conditional independance assumptions in the calculation of correlations of relatives e) linear relationships in regression of phenotypes valueson an individual or relative; f)the independance of gene environmental interactions; g) the assump-tion of constant within sibship variance h) gaussian distribution of phenotypes i) A male and female x and y are joined by a preference (selection) process that is intrinsically non linear. »
- Ma traduction de ce dernier point : L'amour est intrinsèquement non linéaire !

Conclusion (1)

- Nous avons resitué le texte de Fisher 1918 dans le contexte plus large de son œuvre statistique entre 1918 et 1930
- Le biais de cet exposé est dans notre choix de privilégier la filière statistique aux dépends des considérations eugénistes et génétiques qui seront développées par les deux orateurs suivants. Mais ce biais a permis je l'espère de mieux évaluer
  - ce que Fisher a du dépasser pour unifier statistique et génétique avant de les « réconcilier »
  - Ce qu'ont apporté ses concepts statistiques (logique inductive vraisemblance, estimation, analyse de variance), et ses concepts génétiques (héritabilité) à cette œuvre de synthèse
  - On ne peut pas cependant parler de « réconciliation »
  - La petite excursion en économétrie a montré que les problèmes de réinterprétation des mathématiques de la théorie des erreurs ne sont pas propres au domaine de la génétique

Conclusion (2)

- La difficulté face au texte de Fisher n'est pas seulement dans les 3 disciplines qu'il articule et dans les nouveaux concepts qu'il propose. Elle est aussi dans sa manière d'argumenter mathématiquement avec beaucoup d'intuitions et de raccourcis. La syntaxe de Fisher est jugée très obscure » (Bowley), « inaccessible à une bonne part des lecteurs » (Isserlis) et « antipédagogique » (Greenwood qui en dénonce le charabia). Les courts- circuits de sa pensée obligent le lecteur à construire lui-même les ponts qui enjambent les gaps multiples de son raisonnement.
- Fisher défend cette prédilection pour **une mathématique qui ne se réduit pas à un jeu syntaxique** : « la rigueur telle qu'elle est comprise en mathématiques déductive n'est pas suffisante. Dans le raisonnement déductif, des conclusions fondées sur un petit nombre de postulats acceptés n'ont besoin que de rigueur mathématique pour garantir leur vérité (…) Le raisonnement inductif ne peut pas prétendre à une vérité qui est moins que toute la vérité. (…) Ceci est un avertissement à ceux qui peuvent être tentés de penser que le code particulièrement précis des énoncés mathématiques auquel ils ont été exercé au College est est un substitut à l'usage de forces de raisonnement que l'humanité a probablement possédé depuis les temps préhistoriques, et pour lesquelles le processus de codification est encore incomplet, comme le montre l'histoire de la théorie des probabilités. »