

PONTIFÍCIA UNIVERSIDADE CATÓLICA DE SÃO PAULO
PUC-SP

Katia Regina Venturineli

O princípio da homotipose em Karl Pearson:
aspectos biológicos e estatísticos (1899 – 1906)

MESTRADO EM HISTÓRIA DA CIÊNCIA

SÃO PAULO

2010

PONTIFÍCIA UNIVERSIDADE CATÓLICA DE SÃO PAULO
PUC-SP

Katia Regina Venturinel

O princípio da homotipose em Karl Pearson:
aspectos biológicos e estatísticos (1899 – 1906)

MESTRADO EM HISTÓRIA DA CIÊNCIA

Dissertação apresentada à
banca examinadora da
Pontifícia Universidade Católica
de São Paulo, como exigência
parcial para obtenção do título
de MESTRE em História da
Ciência, sob a orientação da
Profa. Dra. Lilian Al-Chueyr
Pereira Martins.

SÃO PAULO

2010

Venturineli, Katia Regina

“O princípio da homotipose em Karl Pearson:
aspectos biológicos e estatísticos (1899-1906)”

São Paulo, 2009
xi, 70 p.

Dissertação (Mestrado) – PUC- SP

Programa: História da Ciência

Orientadora: Profa. Dra. Lilian Al-Chueyr Pereira Martins.

Banca Examinadora

Autorizo, exclusivamente para fins acadêmicos e científicos, a reprodução total ou parcial desta dissertação por processos fotocopiadores ou eletrônicos.

Ass.: _____

Local e data: _____

Katia Regina Venturineli

caixapostalkatia@hotmail.com

Procure a sabedoria e aprenda a escrever os capítulos mais importantes de sua história nos momentos mais difíceis de sua vida.

Augusto Cury

Dedicatória

À minha querida orientadora Professora Lilian AL-Chueyr Pereira Martins pela paciência, amizade, compreensão e apoio aos momentos mais difíceis.

Ao meu marido, André, por sua paciência e carinho.

À minha querida irmã, Mônica.

Aos meus pais, Pedro e Eugênia.

AGRADECIMENTOS

A todos os professores do mestrado, especialmente Professora Maria Elice Brzezinki Prestes, Professor Roberto de Andrade Martins e Professor Ubiratan D'Ambrósio, por suas contribuições.

A todos os colegas do mestrado pelo companheirismo e generosidade.

.

RESUMO

Karl Pearson (1857-1936) é geralmente conhecido por suas contribuições relacionadas à matemática e estatística. Entretanto, além disso, ele apresentou (em 1901 e 1902) uma proposta em relação à biologia: o princípio da homotipose. Para testá-lo, ele aplicou a estatística. A proposta de Pearson gerou uma discussão com William Bateson (1861-1926) e outros biólogos da época.

O objetivo desta dissertação é inicialmente descrever a proposta de Pearson relativa ao princípio da homotipose tanto em termos biológicos como em termos estatísticos e, a seguir, analisá-la. Procurará averiguar se a recepção negativa por parte dos mendelianos e outros biólogos se deveu a uma fundamentação fraca.

Esta dissertação contém uma introdução e quatro capítulos. O capítulo 1 discute alguns precedentes dos estudos estatísticos de Karl Pearson. O capítulo 2 trata da carreira e interesses profissionais abordando os aspectos biológicos relacionados à homotipose. O capítulo 3 analisa os conceitos estatísticos de Pearson e as correlações verificadas em seus estudos sobre homotipose. O capítulo 4 apresenta algumas considerações finais sobre o que foi discutido nos capítulos anteriores.

Este estudo mostrou que na parte relacionada à estatística Pearson empregou de modo adequado métodos como coleta, contagem, tabulação, cálculo e comparação de dados estudados. Apesar disso, a concepção da homotipose, seu ponto de partida, em termos biológicos, era equivocada.

Palavras-chave: História da biologia; História da estatística; Pearson, Karl; homotipose.

ABSTRACT

Karl Pearson (1857-1937) is usually known for his contribution to mathematics and statistics. Besides, he presented (1901-1902) a proposal to Biology: The homotyposis principle. He applied the statistics to test it. Pearson's proposal generated an argument with William Bateson (1861-1926) and other biologists.

This dissertation's aim is, firstly, to describe Pearson's proposal in relation to homotyposis in accordance with both biology and statistics terms, and then, to analyse it. It should also check if the negative reception by the Mendelians and other biologists was due to a weak basis.

This dissertation contains an introduction and four chapters. Chapter 1 discusses some Karl Pearson's previous statistical studies. Chapter 2 is about his career and professional interests dealing with biological aspects in relation to homotyposis. Chapter 3 analyses Pearson's statistical concepts and its correlation verified in his homotyposis studies. Chapter 4 presents some conclusions about what was discussed in the previous chapters.

This study showed that Pearson applied adequate methods accurately regarding statistics, such as collection, counting, tabulation, calculation and comparison of data according to the data studied. Nevertheless, the conception of homotyposis, its starting point in biological terms, was mistaken.

Key-words: The history of biology; the history of statistics; Pearson, Karl; homotyposis.

SUMÁRIO

INTRODUÇÃO	1
CAPÍTULO 1 - ALGUNS PRECEDENTES DOS ESTUDOS ESTATÍSTICOS DE KARL PEARSON	4
1.1 Algumas contribuições matemáticas e estatísticas nos séculos XVIII e XIX	5
1.2 Adolphe Quetelet e suas aplicações da curva normal	9
1.3 Francis Galton e a distribuição normal através do “quincunx”	10
1.4 Karl Pearson e suas influências estatísticas.....	17
CAPÍTULO 2 - KARL PEARSON E OS CONCEITOS BIOLÓGICOS	20
2.1 Karl Pearson: carreira e contexto.....	20
2.2 Mendelianos versus biometricistas	30
2.3 Conceitos biológicos de Karl Pearson e o princípio da homotipose no início do século XX.....	35
CAPÍTULO 3 - ANÁLISE MATEMÁTICA DO TRABALHO DE KARL PEARSON SOBRE HOMOTIPOSE.....	42
3.1 Conceitos estatísticos e a correlação	42
3.2 Aspectos estatísticos do princípio da homotipose	46
3.3 Resultados da correlação	52
CAPÍTULO 4 - CONSIDERAÇÕES FINAIS.....	62
ANEXO	65
BIBLIOGRAFIA.....	68

INTRODUÇÃO

O matemático e estatístico inglês Karl Pearson (1857-1936), personagem principal deste estudo, deixou contribuições na vasta área da estatística. Aplicou os estudos estatísticos à biologia e se considerava um seguidor de Charles R. Darwin (1809-1882). Considerando o grande volume de publicações e estudos deixados por Pearson, decidimos focar essa dissertação acerca dos aspectos biológicos e estatísticos ao estudo do princípio da homotipose, proposto por ele. Seu artigo sobre o princípio da homotipose “Mathematical Contributions to the Theory of Evolution IX” em 1901 provocou muitas reações por parte dos biólogos da época, principalmente aqueles que seguiam uma linha de pesquisa mendeliana. O período selecionado para essa pesquisa vai de 1899 (quando aparecem os primeiros estudos sobre homotipose) até 1906 (ocasião em que houve uma mudança significativa na vida de Pearson, causada pela morte de seu amigo Weldon).

Além da publicação de diversos trabalhos na área da estatística aplicada, Pearson junto com Walter Frank Raphael Weldon (1860-1906) desenvolveram a escola biometricista, fundada por Francis Galton (1822-1911). Esta se caracterizava pela aplicação da estatística aos problemas biológicos voltados a hereditariedade.

Em seu artigo sobre a homotipose, Pearson propôs o conceito biológico de que um organismo individual produz “órgãos similares indiferenciados”, chamando-os de homotipos e que a homotipose seria “o princípio que os homotipos são correlacionados”¹. Acrescentou que a hereditariedade é somente um caso especial da homotipose. Este conceito provocou discussões com os biólogos. Os mendelianos liderados por Willian Bateson discordavam em vários aspectos dos conceitos sobre hereditariedade adotados pelos biométricos, inclusive sobre a homotipose.

¹ Karl Pearson, “Mathematical Contributions to the Theory of Evolution. IX. On the Principle of Homotyposis and Its Relation to Heredity, to the Variability of the Individual, and to that of the Race. Part I. Homotypos in the Vegetable Kingdom”, *Philosophical Transactions of the Royal Society of London* 197 (1901):285-379, na p.294.

Levamos em consideração algumas contribuições de matemáticos que precederam a Pearson, que trabalharam o desenvolvimento do cálculo de probabilidade e métodos estatísticos. Percebemos que antes do século XIX, os cálculos estatísticos não se aplicavam às ciências biológicas. Isso aconteceu somente quando a análise estatística de Adolphe Quetelet (1796-1874) aplicou-se aos dados demográficos e sociais, fazendo o uso particular da distribuição normal que tinha originado como uma aproximação à distribuição binomial no trabalho de Abraham De Moivre (1667- 1754) e tinha se tornado rapidamente o centro da teoria estatística. Neste cenário, o matemático e astrônomo belga Quetelet, desempenhou um papel importante no desenvolvimento das estatísticas na Europa. Entretanto, foi a contribuição do inglês Francis Galton que afetou mais o desenvolvimento teórico. No século XIX, Galton aplicou a distribuição normal ao fenômeno biológico não somente com a curva de erros, mas também com a curva de variação, aplicadas a alguns conceitos da teoria da evolução de Charles Darwin. Seu objetivo era mensurar, estudar, fracionar e correlacionar tal variabilidade biológica. Nesta época Pearson defendia que a distribuição normal não era a única distribuição provável para a análise e interpretação estatística dos dados, dentre outros estudos estatísticos.

Nessa pesquisa apresentaremos algumas concepções estatísticas de Karl Pearson, no período considerado (1899-1906), relacionadas ao princípio da homotipose, além de sua coleta de materiais e tabulações presentes em seu artigo, verificaremos as concepções biológicas na qual ele se baseou para a elaboração de seu princípio.

Nesse estudo utilizaremos as fontes primárias de Karl Pearson, como livros, artigos e cartas. Também examinamos outras obras originais de outros autores como Galton, Bateson, e fontes secundárias tais como os livros e artigos de Stigler, Provine, Eisenhart, Magnello, Martins, Porter, por exemplo, e outros artigos relacionados ao assunto estudado.

Esta dissertação contém uma introdução e quatro capítulos. O capítulo 1 discute alguns precedentes dos estudos estatísticos de Karl Pearson. O capítulo 2 trata da carreira e interesses profissionais abordando os aspectos biológicos relacionados à homotipose. O capítulo 3 analisa os conceitos estatísticos de Pearson e as correlações verificadas em seus estudos sobre

homotipose. O capítulo 4 apresenta algumas considerações finais sobre o que foi discutido nos capítulos anteriores.

CAPÍTULO 1

ALGUNS PRECEDENTES DOS ESTUDOS ESTATÍSTICOS DE KARL PEARSON

No século XIX, no continente europeu, considerava-se de um modo geral, os estudos nas áreas de estatística social e antropométrica como um ramo tradicional da aplicação da teoria das probabilidades. Na Inglaterra, ao contrário, a estatística desenvolveu-se através da aplicação à biologia. Um dos nomes relevantes que deve ser mencionado nesta conexão é Francis Galton (1822-1911)². Galton foi o fundador da escola biometricista. Destacou-se por propor a utilização de ferramentas estatísticas (regressão e correlação) para o estudo da hereditariedade e pela aplicação de métodos estatísticos ao estudo da evolução³. Algumas das concepções de Galton foram desenvolvidas e modificadas pelo biólogo Walter Frank Raphael Weldon (1860-1906) e pelo matemático e estatístico Karl Pearson (1857-1936), seus seguidores.

A escola biometricista caracterizou-se por aplicar métodos estatísticos aos estudos biológicos. Na época, de um modo geral, aqueles que se dedicavam aos estudos biológicos não tinham uma formação mais aprofundada em matemática e nem aplicavam métodos estatísticos aos mesmos como os biometricistas passaram a fazer. As divergências conceituais, metodológicas, dentre outras, levaram os biometricistas ou ancestralistas a entrarem em conflito como os biólogos mendelianos, liderados por William Bateson, em uma acirrada controvérsia que ocorreu entre 1902 e 1906 na Grã Bretanha, mas que teve suas raízes em episódios anteriores⁴. Um deles foi justamente a discussão

² Andrei Nikolaevich Kolmogorov & Adolf-Andrei Pavlovich Yushkevich, *Mathematics of the 19th century*. (Berlin: Birkhauser, 1992), p. 272.

³ Andreza Polizello, *Modelos microscópicos de herança no século XIX: a teoria das estirpes de Francis Galton*. [Dissertação de Mestrado]. São Paulo: PUC, 2009, p. 1.

⁴ Lilian Al-Chueyr Pereira Martins, "Pearson, Bateson e a controvérsia mendeliano-biometricista: uma disputa entre evolucionistas", *Filosofia Unisinos* 8 (2, 2007): 170-190, na p. 176.

em periódicos científicos que se travou entre Pearson e William Bateson em relação ao princípio da homotipose⁵.

Entre 1890 e 1910, ocorreram várias mudanças estruturais em quase todos os campos da atividade humana, da economia à história e à música, e a matemática não foi uma exceção. Os principais fatores desta transformação foram as investigações relacionadas aos fundamentos das matemáticas e o desenvolvimento de estruturas abstratas na álgebra, lógica e espaços gerais. Houve um desenvolvimento em outras áreas como a física matemática (através da teoria quântica e da relatividade), na biologia (biométrica) e na engenharia⁶. Segundo Eileen Magnello, nessa época, Karl Pearson forneceu aos cientistas um conjunto de ferramentas quantitativas para conduzir a pesquisa, acompanhada de uma linguagem científica universal que padronizou a escrita científica no século XX.⁷

Este capítulo apresentará alguns conceitos estatísticos e matemáticos vigentes durante o século XIX, tanto no período em que Karl Pearson deu suas contribuições em relação ao princípio da homotipose (1899-1906), objeto de estudo desta pesquisa, como no período anterior. Nesse sentido, voltaremos um pouco no tempo para recuperar alguns aspectos importantes.

1.1 ALGUMAS CONTRIBUIÇÕES MATEMÁTICAS E ESTATÍSTICAS NOS SÉCULOS XVIII E XIX

No século XVIII, houve um importante achado para os estudos estatísticos: um raro apêndice de um volume do matemático francês Abraham De Moivre (1667-1754). Nesta obra, aparece uma curva que representa a distribuição normal de fenômenos, reproduzida na segunda edição de seu *The Doctrine of Chances* (1738). O autor trabalhava a aproximação de distribuições binomiais para grandes valores de n . Segundo Burton H. Camp, este achado

⁵ Lilian Al-Chueyr Pereira Martins, “Karl Pearson, William Bateson e a controvérsia da homotipose”, *Episteme* (2006):1-9, na p.1.

⁶ Dirk J. Struik, *História concisa das Matemáticas*. Trad. João Cosme Santos Guerreiro. 2ª edição. (Lisboa: Ed. Gradiva, 1992), p. 304.

⁷ Eileen Magnello. “Karl Pearson and the Origins of Modern Statistics: An Elastician becomes a Statistician”. *The Rutheford journal* 2 (2006-2007):1-12, na p.1.

mostra que ao contrário do que alguns pensavam, a distribuição normal é uma ideia original de De Moivre e não de outros matemáticos como Johann Carl Friedrich Gauss (1777-1855) e Pierre Simon Laplace (1749 -1827).⁸

Em 1733, Abraham de Moivre desenvolveu a equação matemática da curva normal, que serviu de base para parte da teoria da estatística indutiva. Curva normal não é o único nome que tem sido dado a esta curva. Também referimo-nos à distribuição normal como *distribuição gaussiana*, em homenagem ao matemático, astrônomo e físico alemão Johann Carl Friedrich Gauss que estudou a curva em 1809 e derivou sua equação a partir de um estudo dos erros em medições repetidas da mesma quantidade⁹ porém, não se pode dizer que a tenha introduzido.

O termo “normal”, utilizado como descrição da curva de probabilidade, apareceu independentemente na obra do matemático americano Charles Sanders Peirce (1839-1914) em 1873, na obra do inglês Galton em 1877, e também na obra do estatístico alemão Wilhelm Lexis (1837-1914) em 1877. Assim, essa ideia foi difundida na década de 1870 em três países e dois idiomas, sendo aplicada à população de pessoas, medidas e suas semelhanças.¹⁰

Recentemente, temos a tendência em nos referir à curva como a curva com formato de sino, terminologia empregada por Esprit Jouffret que utilizou o termo “superfície de sino” (em 1872, p. 222; em 1875, p. 5)¹¹. Também encontramos o nome “padrão”, proposto por Augustus De Morgan (1806-1871) em 1838. Este autor mencionou que poderia ser chamada no futuro de “lei padrão da facilidade do erro”¹². Porém, não seguiu essa sugestão em seus trabalhos posteriores.¹³

⁸Burton H. Camp, “Karl Pearson and Mathematical Statistics”, *Journal of The American Statistical Association* (184, 1933): p.395-401, na p. 399.

⁹ Ronald E. Walpole [ET AL.], *Probabilidade e estatística para engenharia e ciências*. (São Paulo: Pearson Prentice Hall, 2009), pp. 112-113.

¹⁰ Stephen M. Stigler, *Statistics on the table: The history of statistics concepts and methods* (Cambridge, MA: Harvard University Press, 2000), p. 404.

¹¹ Esprit Jouffret, “Étude sur l’effet utile du tir” . *Revue maritime et coloniale* 33 (1872) :205 – 246, na p.222 ; Esprit Jouffret, *Sur la probabilité du tir des bouches à feu et la méthode des moindres carrés*. (Paris :Ch. Tanera, 1875) p.5.

¹² Augustus De Morgan, *An Essay on Probabilities, and on Their Application to Life Contingencies and Insurance Offices*. (London: Longman, Orme, Brown, Green & Longmans, and John Taylor, 1838), p.143.

¹³ Stephen M. Stigler. *Statistics on the table: The history of statistics concepts and methods*. (Cambridge, MA: Harvard University Press, 2000), p. 404.

O gráfico da distribuição normal, também é conhecido como “curva normal dos erros”; é utilizado para descrever diversos fenômenos que ocorrem na natureza e nas pesquisas¹⁴ e apresenta-se como:

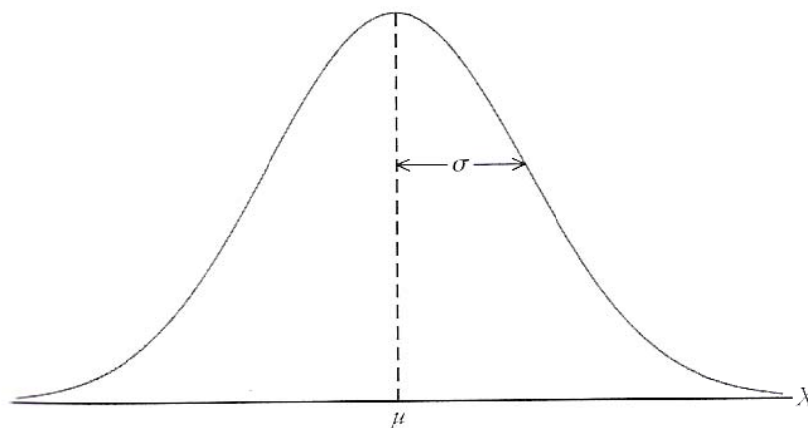


Figura 1.1: A curva normal.

(Fonte: Ronald E. Walpole [et al.], *Probabilidade e estatística para engenharia e ciência*, p.113)

O gráfico da distribuição normal, além de ser simétrico, prolonga-se indefinidamente em ambas as direções, sendo uma curva assintótica, ou seja, ela se aproxima cada vez mais do eixo horizontal sem jamais tocá-lo. Outra característica importante da distribuição normal é que ela depende de dois parâmetros, μ e σ , respectivamente sua média e seu desvio-padrão. Essa curva é chamada de *distribuição normal padronizada* ou *padrão*, quando os valores de média e desvio-padrão correspondem a $\mu=0$ e $\sigma=1$.¹⁵ A função de densidade para esta distribuição é dada por:¹⁶

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$$

a equação matemática deste tipo de curva tem $-\infty < x < \infty$, onde e é o número irracional 2,71828...

¹⁴ Ronald E. Walpole [et al.], *Probabilidade e estatística para engenharia e ciências*. (São Paulo: Pearson Prentice Hall, 2009), p. 112.

¹⁵ Gary A. Simon, *Estatística aplicada*. (Porto Alegre: Atlas, 2000), pp. 164 - 165.

¹⁶ Murray R. Spiegel. *Probabilidade e estatística*. (São Paulo: Pearson, 1978), p.157.

Karl Pearson foi quem introduziu o termo “desvio-padrão” em 1893 e com o tempo mudou a noção do que era distribuição normal “padrão”. Pearson também relatou suas reivindicações sobre o termo “normal” como seu, cujo uso no início de 1893, foi seguido por passagens de 1904 a 1906 e em 1920.¹⁷ Pearson relatou que seu costume em chamar a curva Gauss-Laplaciana de curva normal lhe poupava de atribuir o mérito da descoberta a um dentre dois grandes astrônomos matemáticos. Pearson comentou:

Muitos anos atrás eu chamei a curva Laplace-gaussiana de curva normal, cujo nome, enquanto evita uma questão internacional de prioridade, tem a desvantagem de levar as pessoas a acreditarem que todas as outras distribuições de frequência, são de um modo ou de outro, “anormais”.¹⁸

Pearson já se referia à curva de probabilidade normal em suas publicações anteriores¹⁹. Alguns estudos historiográficos mencionam que Henri Poincaré (1854 —1912) teria empregado o termo em 1893-1894: “ Para ser breve, direi que a lei da probabilidade é normal quando o valor da probabilidade é representado por este integral.”²⁰ Tanto Poincaré quanto Pearson não mencionaram autores que tivessem utilizado a curva de probabilidade normal anteriormente. Estudiosos, como Victor Hils²¹, dão a prioridade a Francis Galton em seu *Natural Inheritance* em 1889.²²

A distribuição normal, em sua formulação por Gauss e depois por Laplace, representa a distribuição dos erros. Durante a primeira metade do século XIX, o matemático belga Lambert Adolphe Jacques Quetelet (1796-

¹⁷ Stephen M. Stigler, *Statistics on the table: The history of statistics concepts and methods* (Cambridge, MA: Harvard University Press, 2000), pp. 405-406.

¹⁸ Karl Pearson, “Notes on the history of correlation”, *Biometrika* 13 (1920): 25 - 45, na p. 25.

¹⁹ Stephen M. Stigler, *Statistics on the table: The history of statistics concepts and methods*, p. 406.

²⁰ Henri Poincaré, *Calcul des probabilités*, (Paris : Gauthier-Villars, 1896), p. 76.

²¹ Ver por exemplo: Victor Hils, “Statistics and Social Science,” in R. Giere and R. Westfall (eds.), *Foundations of Method: The Nineteenth Century* (Bloomington: Indiana University Press, 1973), pp. 206-233.

²² Stephen M. Stigler, *Statistics on the table: The history of statistics concepts and methods*, p. 407.

1874) dela se apropriou²³, desenvolvendo seus estudos como veremos a seguir.

1.2 ADOLPHE QUETELET E SUAS APLICAÇÕES DA CURVA NORMAL

O estatístico belga Adolphe Quetelet (1796-1874) trouxe contribuições para a análise estatística dos dados sociais, tais como o conceito do homem médio e o ajustamento da distribuição normal²⁴, junto com a interpretação da regularidade estatística. Em sua principal obra, *Essai de Physique Sociale*, publicada em 1835, ele analisou as qualidades físicas, morais e intelectuais do homem; as propriedades do homem médio e o sistema social. Quetelet tomou conhecimento da curva normal, que foi fundamental para o desenvolvimento do conceito do homem médio.²⁵

Quetelet não somente trabalhou com características físicas como altura e peso, mas também com características morais como as tendências dos indivíduos em cometer crimes ou se tornarem alcoólatras. Ele notou que muitas das características que havia coletado poderiam fazer parte da curva normal, isto é, havia um valor principal e “erros” do meio que estavam sendo distribuídos da mesma maneira que os erros de medida.²⁶

Em 1846, Quetelet escreveu uma carta ao Grand Duke de Saxe Coburg, onde se mostrou favorável à utilização da curva normal. Quetelet sugeriu que se partisse da suposição de que, se alguém fizesse mil cópias de uma estátua, essas cópias estariam naturalmente sujeitas a uma grande variedade de erros, que iriam se combinar num simples padrão de erros. Quetelet fez o experimento e suas cópias de estátuas eram seres vivos. As estátuas de Quetelet eram soldados escoceses: ele mediu a circunferência do peito de 5.732 soldados e notou que elas eram distribuídas normalmente por cerca de

²³ Victor J. Katz. *A History of Mathematics: an introduction*. (Reading, MA: Addison-Wesley, 1998), p.758.

²⁴ Quetelet além de ter calculado as médias aritméticas das medidas, originalmente, considerou suas dispersões e descobriu que a curva normal, conhecida na época como a curva dos erros, podia ser ajustada satisfatoriamente às medidas de peso, estatura e perímetro torácico.

²⁵ Victor J. Katz. *A History of Mathematics: an introduction*. (Reading, MA: Addison-Wesley, 1998), p. 758.

²⁶ *Ibid.*

mais ou menos 40 polegadas. Sua conclusão foi que as medidas foram distribuídas como se fossem de uma natureza objetivando um tipo ideal. Sua distribuição mostrou que havia o soldado escocês comum e o desvio de “comum” foi simplesmente devido à combinação de causas acidentais.²⁷

Segundo Victor Katz, nessa época, muitos discordaram do programa de Quetelet que defendia a busca de uma distribuição normal para todas as situações. Quetelet deu muita importância à curva normal, devido à sua crença no determinismo²⁸. A ideia da distribuição normal tornou-se o centro de muitas discussões envolvendo estatística. Acreditava-se na época na existência de um valor estatístico ideal do meio e que a curva normal era a curva ideal, desde que se seguisse a lei dos erros. Por isso, toda variação do meio deveria seguir a curva.²⁹

Francis Galton utilizou as ideias de Quetelet, procurando matematizar a teoria da evolução de Darwin considerando a variação da hereditariedade. Galton fez vários experimentos envolvendo tanto vegetais como animais. Ele estudou a variação do tamanho de ervilhas e a estatura dos descendentes em relação à estatura de seus pais. Ele focou o estudo estatístico relacionado à distribuição normal, variabilidade e correlação, assunto que discutiremos a seguir.

1.3 FRANCIS GALTON E A DISTRIBUIÇÃO NORMAL ATRAVÉS DO “QUINCUNX”

Nos anos de 1874 a 1877, Galton direcionou seus estudos às qualidades mensuráveis, pois, ao abordar fenômenos referentes à hereditariedade do gênio humano (assunto que muito lhe interessava), verificou que essa qualidade intelectual não permitia medidas em grande escala. Galton voltou-se então, a qualidades mensuráveis como a estatura em humanos e também

²⁷ Victor J. Katz. *A History of Mathematics: an introduction*, p.758.

²⁸ A ideia de que todo dado tinha que confrontar a distribuição normal foi enraizada nas ideologias filosóficas de determinismo e essencialismo de Aristóteles, ideia que se estendeu até o final do século XIX.

²⁹ Eileen Magnello, “Karl Pearson and the Origins of Modern Statistics: An Elastician becomes a Statistician”. *The Rutheford Journal* 2 (2006-2007):1-12, na p. 2.

iniciou uma série de experiências envolvendo medições em gerações sucessivas do diâmetro das ervilhas (doces).³⁰

Galton realizou vários experimentos. Em um deles, feito em 1875, ele estudou o diâmetro das sementes de um tipo particular de ervilha³¹. Escolheu certa quantidade de sementes de ervilhas com sete tamanhos diferentes, cruzando-as. Ao analisar seus descendentes, percebeu que, na progênie de cada grupo, os tamanhos eram normalmente distribuídos, mas que a variabilidade em cada grupo - isto é, a distribuição de dados - foi essencialmente a mesma. Em suas pesquisas sobre o diâmetro de ervilhas, chegou a uma equação, por ele denominada de reversão à média:

$$y - \bar{y} = r (x - \bar{x})$$

onde os y exprimem o diâmetro das sementes das plantas-filhas e os x o das plantas-pais; r significa reversão. No trabalho de Galton "Typical Laws of Heredity in Man", aparece pela primeira vez uma medida numérica r que foi chamada de reversão e que Galton mais tarde chamou de regressão.³² Segundo Pearson, "este r é a fonte do nosso símbolo para o coeficiente de correlação, que foi na realidade a primeira letra de reversão e não de regressão."³³

Galton tabulou esses dados e mapeou os pontos, estimando a inclinação da reta que melhor conectaria esses pontos.³⁴ Ele desenvolveu vários outros

³⁰ Stephen M. Stigler. *Statistics on the table: The history of statistics concepts and methods*, p. 177.

³¹ Galton verificou que era muito difícil coletar material humano para duas gerações e depois de cuidadosa consideração selecionou sementes de ervilha doce.

³² Durante experimentos envolvendo sementes de ervilha que tinham por objetivo determinar a lei da herança do tamanho, empregou cem sementes-pai de cada um dos sete tamanhos diferentes escolhidos, ele construiu um gráfico bidimensional com os diâmetros das sementes-pai e das sementes descendentes de cada classe parental. Notou que os diâmetros medianos dos descendentes das classes parentais contendo as sementes maiores eram menores do que os de seus pais, e para as classes contendo as sementes menores, eles eram maiores, o que parecia indicar uma tendência do tamanho "médio" de semente de "reverter" em direção ao que poderia ser descrito como o tipo ancestral médio. Em seus estudos, Galton, também verificou que, quando os pais eram mais altos do que a média, os filhos tendiam a ser menores do que eles e, quando os pais eram mais baixos do que a média, os filhos tendiam a ser maiores do que eles. Verificou que as oscilações para mais ou para menos ocorriam dentro de certos limites, tendo a média aritmética da espécie como ponto de convergência.

³³ Karl Pearson, "Notes on the History of Correlation", *Biometrika* 13 (1, 1920):25-45, na p. 33.

³⁴ Victor J. Katz. *A History of Mathematics: an introduction*. (Reading, MA: Addison-Wesley, 1998), p. 759.

estudos envolvendo a hereditariedade, incluindo um extenso trabalho sobre a estatura das crianças em relação à estatura de seus pais e encontrou resultados semelhantes aos que havia encontrado nos experimentos com as ervilhas.

Galton compreendeu que, em muitas circunstâncias, as duas variáveis que ele estava considerando estavam “correlacionadas”, isto é, a inclinação das linhas de regressão de uma sobre a outra eram as mesmas quando ambos os grupos de dados eram escalados pelas unidades do erro provável. Segundo Katz, Galton compreendeu que uma forte correlação não implicava necessariamente em um relacionamento causal.³⁵

Enquanto desenvolvia essas experiências, Galton concebeu uma máquina, a “Quincunx”.³⁶ Em 1873, ele pediu a um comerciante que construísse o “Quincunx”.³⁷ Esta máquina consistia de uma tábua com um funil no topo, através do qual, grãos de chumbo eram lançados para caírem sobre uma sucessão de linhas de alfinetes, que se reuniam no fundo de um compartimento vertical em pilhas que lembram uma curva normal.³⁸

A figura abaixo, retirada do livro *Natural Inheritance* (1889), mostra uma interpretação esquemática do “Quincunx”.

³⁵ Katz, *A History of Mathematics: an introduction*, p.758.

³⁶ Stephen M. Stigler, *Statistics on the table: The history of statistics concepts and methods*. (Cambridge, MA: Harvard University press, 2000), p. 177.

³⁷ O nome “Quincunx” foi derivado de um modelo semelhante de alfinetes usados em árvores frutíferas, na agricultura inglesa, um modelo que era conhecido como quincunxial, porque sua base era sobre um quadrado de quatro árvores com uma quinta no centro.

³⁸ Stigler, *Statistics on the table: The history of statistics concepts and methods*, p. 178.

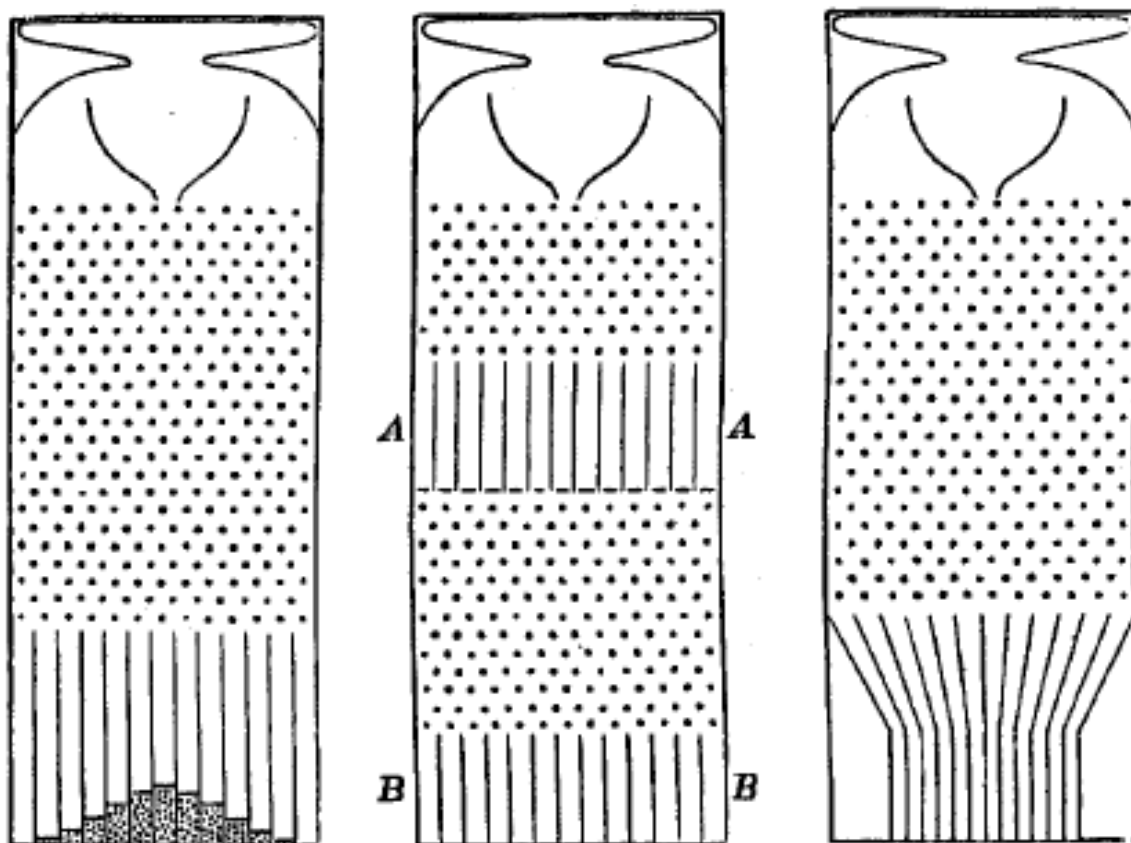


Figura 1.2: Um desenho esquemático do Quincunx.

(Fonte: Francis Galton, *Natural Inheritance*, p. 63)

O “Quincunx” de Galton foi inicialmente destinado a ilustrar o trabalho de um grande número de pequenas causas acidentais que produziam uma distribuição normal.³⁹

Tal experimento fez Galton questionar as grandes causas não-acidentais com as quais havia se deparado em seus estudos. Segundo Stigler, este foi um experimento de pensamento, pois Galton já descrevia o resultado do experimento antes da construção da máquina⁴⁰. Verificamos nas figuras abaixo que antes da construção do Quincunx e a apresentação de suas ideias no *Natural Inheritance*, Galton já apresentara este esquema e suas ideias de curva normal em 1877, em sua obra *Typical Laws of Heredity*:

³⁹ Como os chumbinhos passavam do topo para o fundo, eles eram desviados ao acaso, para cada linha, e, se o aparelho estivesse bem equilibrado, produziriam um traçado no fundo, onde o número de chumbinhos num compartimento seria proporcional ao número de linhas desse compartimento. Isto é, o número de chumbinhos num compartimento seria proporcional aos coeficientes binomiais, quase uma distribuição normal se o número de linhas de alfinetes fosse absolutamente grande.

⁴⁰ Stigler, *Statistics on the table: The history of statistics concepts and methods*, p. 178.

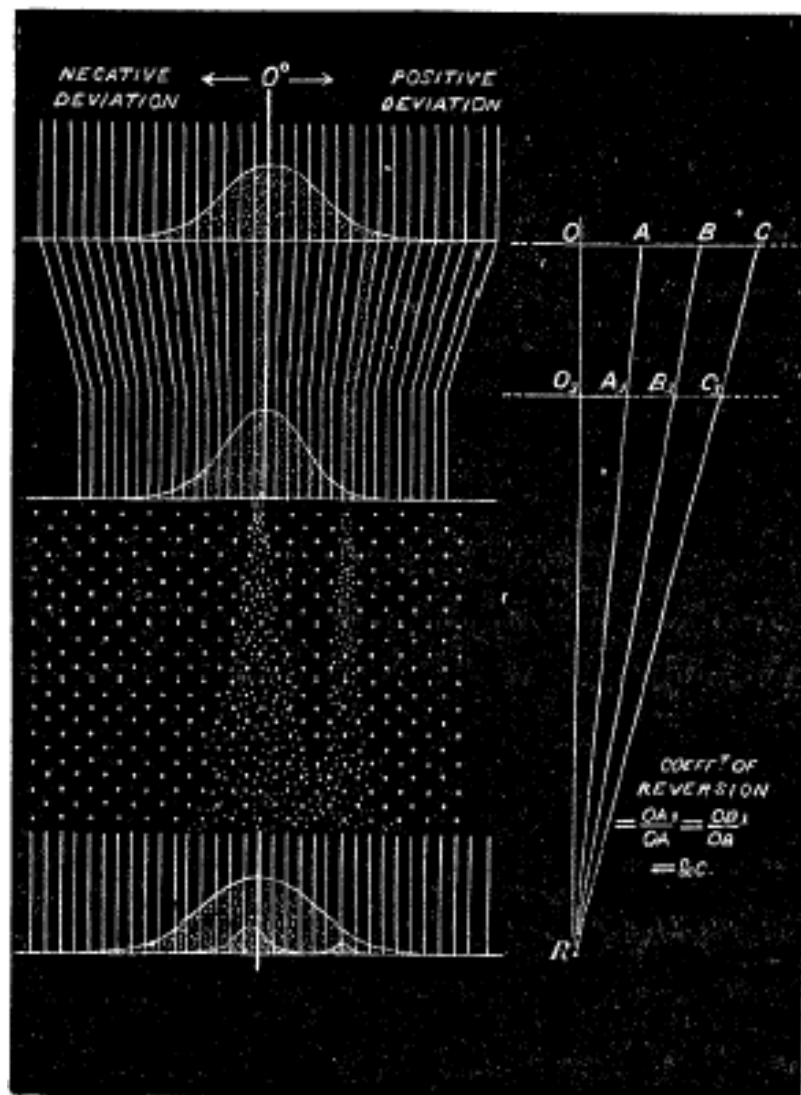


Figura 1.3: Ilustração de Galton, sobre reversão à média.
 (Fonte: Francis Galton, *Typical Laws of Heredity*, p.285)

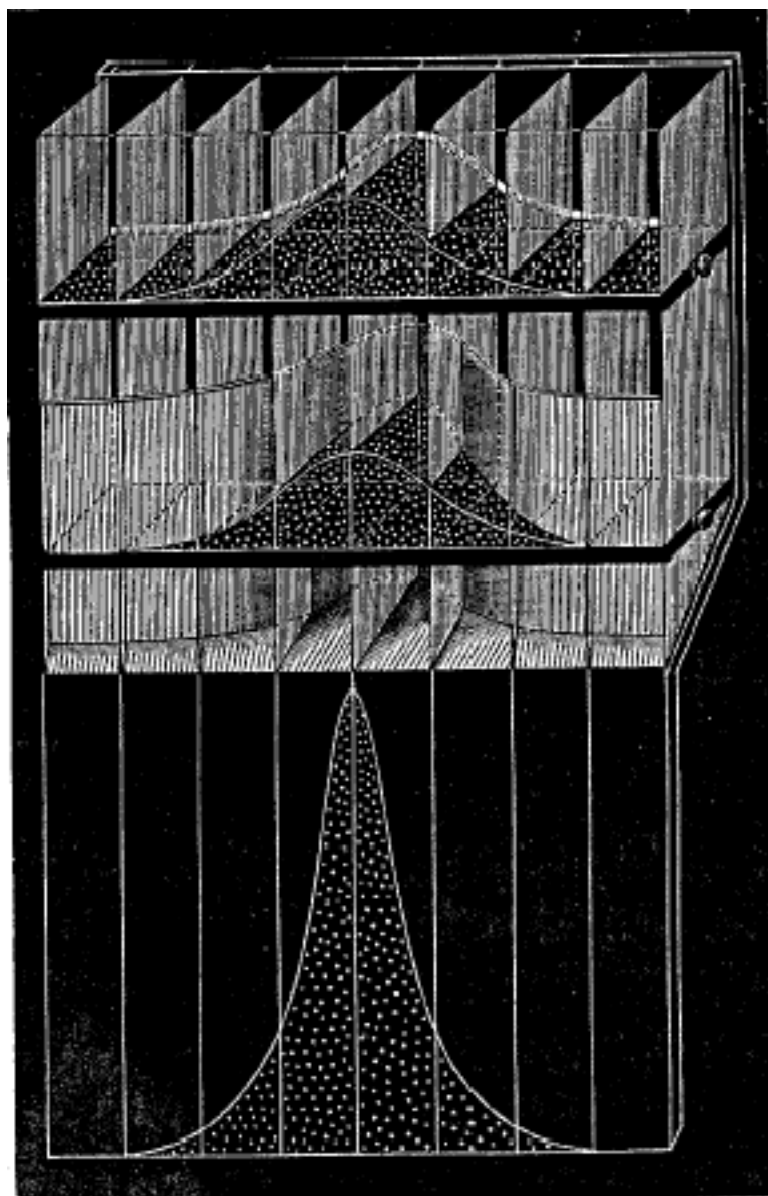


Figura 1.4 : Representação de Galton da curva normal em 1877.

(Fonte: Francis Galton, *Typical Laws of Heredity*, p.297)

Galton, em seu livro *Natural Inheritance*, imaginou alterar o “Quincunx”, desmontá-lo ao meio e esticá-lo, mas isso sem alterar a sua distribuição final de chumbinhos. Para isso, ele acrescentaria barreiras verticais para impedi-los de extraviarem-se, enquanto atravessassem as lacunas. O esquema é mostrado no segundo painel na figura 1.2, no qual Galton ilustrou a ideia, já existente em 1877 (também em correspondência).⁴¹ Com essas barreiras, a introdução das lacunas não teria efeito sobre a distribuição de chumbinhos no

⁴¹ Stephen M. Stigler, *Statistics on the table: The history of statistics concepts and methods*. (Cambridge, MA: Harvard University Press, 2000), pp. 179 - 180.

fundo; Galton, então, introduziu as barreiras na base das lacunas. Tudo isso serviria para reduzir o “Quincunx” e com poucas linhas de alfinetes a serem cruzadas, os chumbinhos ainda permaneceriam numa distribuição normal. Assim, Galton distribuiu os chumbinhos de um nível médio, mas somente em um compartimento: esperava-se produzir uma pequena distribuição normal, imediatamente abaixo do compartimento do qual eles eram distribuídos. Procede-se então, a distribuição aos compartimentos restantes, um de cada vez. Para Galton, cada um produziria a sua própria curva normal, sendo que as mais próximas do centro seriam maiores do que as dos extremos, porque mais chumbinhos seriam depositados nos compartimentos centrais.⁴² A imaginação de Galton mostrou como a curva normal poderia ser separada em componentes.⁴³

Para Galton, o aparelho correspondia às suas investigações sobre hereditariedade, no qual a aparente homogeneidade no final de cada traçado representava uma mistura derivando-se das gerações anteriores. Galton chamou esse fenômeno de “lei do desvio”. Ele acreditava que as causas que atuavam sobre uma característica herdada, tal como a estatura, eram pequenas influências perturbadoras (representadas pelos alfinetes) e que a lei do desvio genético era puramente numérica e seguia universalmente a lei genética da distribuição normal. O pensamento de Galton pode ser interpretado como prova análoga do teorema matemático.⁴⁴ No “Quincunx”, a variável verificada é do tipo discreta, mais especificamente o compartimento no qual o chumbinho irá cair. Nesse caso, emprega-se a distribuição binomial⁴⁵ para descrever o problema.

⁴² Isso no primeiro estágio do Quincunx e quando todos os chumbinhos fossem distribuídos, o resultado, ou seja, a soma de todas as pequenas curvas normais seria como se não tivessem absolutamente existido interrupções.

⁴³ Stephen M. Stigler. *Statistics on the table: The history of statistics concepts and methods*. (Cambridge, MA: Harvard University press, 2000), p. 180.

⁴⁴ *Ibid.*

⁴⁵ A distribuição binomial é usada quando se deseja saber a probabilidade de obter-se x sucessos em n tentativas; se a probabilidade de ter sucesso em uma tentativa é p , consequentemente a probabilidade de se ter fracasso é $q = 1 - p$. Portanto, em cada tentativa ou experimento individual, só dois resultados são possíveis, que são denominados de sucesso ou fracasso, este raciocínio subentendido é binário.

Segundo Magnello, Galton convenceu-se de que todo dado biológico poderia ser normalmente distribuído, e que estava tão comprometido com a ideia da curva que construiu o aparelho “Quincunx”.⁴⁶

Os resultados das experiências de Galton com as ervilhas doces e de seus estudos sobre estatura humana estavam de acordo com os trabalhos iniciais⁴⁷ de Adolphe Quetelet, para o qual o mundo, em geral, era normalmente distribuído.⁴⁸

Segundo Magnello, era tal a tirania da curva normal, que pelo final do século XIX, muitos estatísticos assumiram que nenhuma outra curva poderia ser usada para descrever os dados. Para Magnello, esta visão foi desafiada por Pearson na última década do século XIX.⁴⁹

Outros estatísticos ingleses da época tentaram esclarecer e estender o trabalho de Galton e lutaram contra a filosofia básica anterior ao observar curvas normais representadas em vários tipos de distribuição⁵⁰. Um deles foi Francis Edgeworth (1845-1926) que era muito conservador no uso da curva normal para detectar diferenças significantes.

Em 1892 e 1893, Edgeworth escreveu uma série de artigos sobre correlação, iniciando com *Correlated averages*⁵¹. Estes artigos tomaram como base o trabalho de Galton. Ambos os autores procuravam desenvolver um aparato matemático para a distribuição normal multivariada e para as distribuições condicionais que surgissem em regressões múltiplas. Nestes trabalhos o tratamento foi incompleto. Apesar disso, inspiraram o primeiro trabalho de Karl Pearson sobre correlação.

1.4 KARL PEARSON E SUAS INFLUÊNCIAS ESTATÍSTICAS

⁴⁶ Eileen Magnello, “Karl Pearson and the Origins of Modern Statistics: An Elastician becomes a Statistician”. *The Rutheford Journal* 2(2006-2007):1-12, na p. 2.

⁴⁷ Galton leu os escritos de Quetelet em 1863, dedicando-se à universalidade da distribuição normal.

⁴⁸ Stephen M. Stigler, *Statistics on the table: The history of statistics concepts and methods*. (Cambridge, MA: Harvard University press, 2000), p. 178.

⁴⁹ Eileen Magnello, “Karl Pearson and the Origins of Modern Statistics: An Elastician becomes a Statistician”. *The Rutheford Journal* 2 (2006-2007):1-12, na p. 2.

⁵⁰ Katz, *A History of Mathematics: an introduction*, p. 759.

⁵¹ Para maiores detalhes sobre o trabalho de Edgeworth, ver por exemplo, Francis Ysidro Edgeworth, “Correlated averages”, *Philosophical Magazine* (5th ser.) 34, (1892): 190 - 204.

De acordo com A. M. Kolmogorov, Pearson foi um ávido leitor dos mais remotos mestres da teoria da probabilidade, Jacob Bernoulli (1654-1705), Pierre Simon Laplace (1749-1827), e outros.⁵² Além disso, ele tinha conhecimento dos trabalhos que descrevemos nas seções anteriores e das concepções acerca da estatística na época.

Embora existam diversos estudos historiográficos⁵³ que apontam a influência de Galton sobre Pearson, há outros que não partilham dessa visão.⁵⁴

O trabalho de Galton incluía conceitos e técnicas que ele descobrira e desenvolvera como ferramentas para medir a influência da hereditariedade. Segundo Churchill Eisenhart, com a leitura do *Natural Inheritance*, Pearson ficara encantando com o conceito e implicações da “correlação” de Galton. Assim, começou a construir os fundamentos da ciência estatística a partir de questões propostas por Weldon⁵⁵ que, como mencionamos anteriormente, tinha uma formação biológica. Para Magnello, a interpretação de diversos estudiosos de que a criação de uma nova metodologia estatística por Pearson, está relacionada ao surgimento dos problemas de eugenia, é falsa.⁵⁶ Porém, Pearson expressou claramente sua admiração pela contribuição de Galton em seu trabalho *Notes on the History of Correlation*, ao afirmar: “Eu fiquei imensamente excitado com o livro de Galton de 1889 - *Natural Inheritance* - e eu li um artigo sobre ele no ano do seu surgimento.”⁵⁷ Essa ideia também transparece em várias cartas de Pearson.

Para Churchill Eisenhart, o *Natural Inheritance*, de Galton, convenceu Pearson de que o caminho mais promissor era por meio de estudos estatísticos de variação e correlação nessas populações. Porém, Magnello complementa dizendo que quase todos os historiadores falharam em considerar as ideias de Pearson e Galton, sendo que seus métodos e perspectivas foram diferentes um do outro, assim como duas pessoas o são. Enquanto o foco principal de

⁵² A.N. Kolmogorov, *Mathematics of the 19th century*. (Berlin: Birkhauser, 1992) p. 272.

⁵³ Ver por exemplo, Churchill Eisenhart, “Pearson, Karl”. In: Charles Coulston Gillispie, ed. *Dictionary of scientific biography*. Vol. 9 (New York: Charles Scribner’s sons, 1981). pp 447-473.

⁵⁴ Ver por exemplo Eileen Magnello, “Karl Pearson and the Origins of Modern Statistics: An Elastician becomes a Statistician”. *The Rutheford Journal* volume 2 (2006-2007):1-12.

⁵⁵ Churchill Eisenhart, “Pearson, Karl”. In: Charles Coulston Gillispie, ed. *Dictionary of scientific biography*. Vol. 9 (New York: Charles Scribner’s sons, 1981): 447 - 473, na p. 449.

⁵⁶ Eileen Magnello, “Karl Pearson and the Origins of Modern Statistics: An Elastician becomes a Statistician”. *The Rutheford Journal* 2 (2006-2007):1-12, na p.1

⁵⁷ Karl Pearson, “Notes on the History of Correlation”. *Biometrika* 13 (1, 1920): 25-45, na p. 29.

Pearson foi a exatidão do teste, a ênfase de Galton foi a correlação. Outro fator, para Magnello⁵⁸, foi que Pearson tinha conhecimento de matemática de alto nível, um requisito para fazer estatística, o que tornou seu trabalho matematicamente mais complexo que o de Galton. Pearson interessava-se por grandes grupos de dados (mais de 1000) enquanto Galton concentrava-se mais em pequenos grupos (cerca de 100); Pearson conduziu seu longo projeto durante vários anos, enquanto Galton almejava resultados mais rápidos.

Pearson aplicou seus conceitos e métodos estatísticos em vários trabalhos voltados aos problemas biológicos, contribuindo para o estabelecimento da escola biométrica, juntamente com Galton e Weldon, apresentando diversas publicações nesta área como veremos no capítulo a seguir.

⁵⁸ Magnello, "Karl Pearson and the Origins of Modern Statistics: An Elastician becomes a Statistician", na p.2.

CAPÍTULO 2

KARL PEARSON E OS CONCEITOS BIOLÓGICOS

Neste capítulo descreveremos a carreira do matemático inglês Karl Pearson (1857-1936) seus interesses e estudos. A seguir, comentaremos como Galton, Pearson e Walter Frank Raphael Weldon (1860-1906) deram início à escola biométrica, no final do século XIX, uma escola científica em que se aplicavam aos estudos biológicos métodos estatísticos. Finalmente, apresentaremos as concepções biológicas de Karl Pearson em seu artigo sobre a homotipose, abordando sua análise teórica em relação à correlação entre irmãos com a correlação de “órgãos semelhantes indiferenciados de um indivíduo”.

2.1 KARL PEARSON: CARREIRA E CONTEXTO

Karl Pearson nasceu no dia 27 de março de 1857 em Londres, filho de William Pearson e Fanny Smith, era o segundo de três filhos, sendo o mais jovem dos dois homens. Pearson foi educado em casa até os nove anos de idade, quando foi enviado para a *University College School* de Londres por sete anos, retirando-se em 1873 por problemas de saúde, permanecendo mais um ano com um tutor. Em 1875, obteve uma bolsa de estudos para o *King's College* de Cambridge com um segundo lugar no concurso. Nessa universidade, Pearson estudou matemática com E.J. Routh, G.G. Stokes, J.C. Maxwell, Arthur Cayley e William Burnside.⁵⁹

⁵⁹ Churchill Eisenhart, “Pearson, Karl”. In: Charles Coulston Gillispie, ed. *Dictionary of scientific biography*. Vol. 9 (New York: Charles Scribner's sons, 1981): 447 - 473, na p. 447.

Em 1879, Karl Pearson recebeu o título de bacharel com honras matemáticas pela *Cambridge University*. Após receber o título de bacharel, Pearson viajou para a Alemanha, estudou Física e Metafísica e também assistiu aulas sobre darwinismo. Em 1880, voltou para Londres e fez leituras nas câmaras do Lincoln's Inn e tornou-se *barrister*⁶⁰ em 1881. Pearson recebeu o título de bacharel em Direito pela Universidade de Cambridge e, no ano seguinte o de mestre, porém nunca chegou a praticar tal profissão.⁶¹ No período de 1882 e 1884, Pearson também ministrou em diversos pontos de Londres aulas sobre o pensamento e a vida social na Alemanha desde os primórdios até o séc. XVI, e sobre a influência de Lutero para o bem-estar material e intelectual nessa nação.⁶²

Em 1884, Pearson foi indicado para a cátedra *Goldsmid* de matemática aplicada e mecânica no *University College* de Londres. Durante o período em Cambridge, Pearson não permaneceu restrito à matemática. Também leu Dante, Goethe e Rousseau no original. Antes de deixar a universidade, escreveu resenhas de dois livros sobre Spinoza para a *Cambridge Review* e publicou um artigo sobre Maimônides e Spinoza no periódico *Mind*. Tanto como em Cambridge como na Alemanha, Pearson envolvia-se muito mais do que nas aulas na universidade; interessou-se pelo folclore do país, pela literatura alemã da Idade Média e do Renascimento, e também pelo desenvolvimento de ideias a respeito da posição da mulher.

Pearson tornou-se membro ativo na fundação de um pequeno clube de homens e mulheres, em 1885, dedicado ao debate da relação entre os sexos. Ele ministrou a conferência de abertura, "The Woman's Question" (A questão da mulher), e em outra reunião discursou sobre "Socialism and Sex" (Socialismo e sexo). Entre os participantes encontrava-se Maria Sharpe, com quem se casou em 1890, e teve um filho, Egon, e duas filhas, Sigrid e Helga. Após a morte da esposa, em 1928, tornou a casar-se no ano seguinte, dessa vez com uma colega de departamento, Margaret Victoria Child.⁶³

⁶⁰ Barrister é uma das categorias de advogados que existem na Inglaterra. Sua função principal, mas não exclusiva, é representar os mandatos e os processos ante os tribunais.

⁶¹ Churchill Eisenhart, "Pearson, Karl". In: Charles Coulston Gillispie, ed. *Dictionary of scientific biography*. Vol. 9 (New York: Charles Scribner's sons, 1981): 447-473, na p. 447.

⁶² *Ibid.*

⁶³ *Ibid.*, p. 448.

Através de um contato inicial com Walter Frank Raphael Weldon, orientou os estudos de Pearson, que era um matemático e estatístico, aos aspectos biológicos. Pearson passou a se interessar pelos estudos biológicos. A partir de então eles não apenas passaram a interagir como desenvolveram uma sólida amizade que durou até a morte de Weldon (1906).⁶⁴

Na época, vários estudiosos procuravam seguir as sugestões que Darwin apresentara na *Origin of Species* (no sentido de trazer esclarecimentos sobre a origem das espécies), considerando-se seus seguidores, apesar de muitas vezes interpretarem de modo diferente evolução e hereditariedade. Como menciona Alfred Nordmann, tratava-se de uma “guerra entre darwinistas.”⁶⁵ Embora seguisse vários pressupostos admitidos por Galton, Pearson introduziu⁶⁶ modificações na lei da herança ancestral proposta pelo mesmo.

Entre 1891 e 1894, Pearson lecionou geometria no *Gresham College*, na mesma cidade.⁶⁷ Nessa ocasião, sua única obrigação era ministrar três cursos por ano para ouvintes externos, sobre temas de sua própria escolha. Nos dois primeiros cursos em 1891, Pearson tratou sobre o escopo e os conceitos da ciência moderna, explorando os fundamentos filosóficos da ciência. Estas aulas foram desenvolvidas e expandidas e tornaram-se a primeira edição de *The Grammar of Science* em 1892, livro que influenciou o pensamento científico de toda uma geração. Com essa publicação, Pearson dedicou-se a um projeto de quantificação universal e expôs um conteúdo filosófico, assim como Pearson o chamou, de método científico.⁶⁸ Segundo Pearson, o método científico tem certas características, como a menção cuidadosa e precisa dos dados e observação de sua correlação e sequência. Segundo Porter, *The Grammar of Science* era um tratado filosófico designado a estender o alcance do domínio

⁶⁴ Churchill Eisenhart, “Pearson, Karl”. In: Charles Coulston Gillispie, ed. *Dictionary of scientific biography*. Vol. 9 (New York: Charles Scribner’s sons, 1981): 447- 473, na p. 449.

⁶⁵ Ver por exemplo Alfred Nordmann, “Darwinians at war. Bateson’s place in histories of Darwinism.” *Synthese* 91 (1992): 53-72.

⁶⁶ Os biometricistas aceitavam a ideia de herança com mistura.

⁶⁷ Eisenhart, “Pearson, Karl”, na p. 447.

⁶⁸ A.W.F. Edwards, “R.A. Fisher on Karl Pearson”, *Notes and Records of the Royal Society of London* 48 (1, 1994): 97 - 106, na p. 101.

da ciência, onde Pearson idealizou a estatística como um verdadeiro formulário de todo o conhecimento científico.⁶⁹

Entre novembro de 1891 e maio de 1892, as aulas de Pearson ministradas no *Gresham College* eram dedicadas quase totalmente a um abrangente tratamento formal para a representação gráfica de dados estatísticos das ciências biológicas, físicas e sociais. Continham apenas curtas menções à estatística descritiva numérica. Em *Laws of Chance* (Leis do acaso), entre novembro de 1892 e fevereiro de 1893, Pearson discutia a teoria da probabilidade e o conceito de “correlação”, a partir de experimentos realizados por meio de lançamentos de moeda e de retirada de cartas de um baralho, bem como de observações de fenômenos naturais. O termo “desvio-padrão” foi introduzido em uma aula de 31 de janeiro de 1893, como substituto mais conveniente para o termo “erro da raiz do quadrado da média”.⁷⁰ Em *The Geometry of Chance*, (Geometria do acaso), entre novembro de 1893 e maio de 1894, dedicou uma aula a “curvas normais”⁷¹, outra a “curvas assimétricas” e uma última a “curvas compostas”.⁷² Suas aulas sobre variação e correlação, entre 1892 e 1893, resultaram num material que constituiu os quatro primeiros de seus relatórios sobre evolução publicados em *Philosophical Transactions*. Entre 1893 e 1901, Pearson publicou cerca de trinta e cinco artigos sobre tópicos estatísticos à *Royal Society* e, até 1906, já havia publicado mais de setenta artigos adicionais, contendo outras teorias e aplicações estatísticas.⁷³

Em geral, os estudos de Pearson tinham envolvimento com teorias e aplicações estatísticas, porém suas contribuições a partir de um dado momento passaram a ter o propósito de desenvolver e aplicar métodos estatísticos aos problemas de hereditariedade e evolução, e não desenvolver técnicas tendo como alvo a estatística. Isto fica evidente na apresentação de uma série de publicações que se tornaram conhecidas como “*Contributions to the Mathematical Theory of Evolution*” (1894)⁷⁴. Segundo A. W. F. Edwards, as

⁶⁹ Theodore M. Porter. *The English biometric tradition. Companion Encyclopedia of the History and Philosophy of the Mathematical Sciences*, vol. 2, London: Grattan-Guinness, 1994, p. 1337

⁷⁰ Substituiu também as expressões mais antigas como “erro médio quadrático” e “erro médio”.

⁷¹ Pearson não foi o primeiro a utilizar essa terminologia, Galton também a empregou.

⁷² Eisenhart, “Pearson, Karl”, na p.450.

⁷³ *Ibid.*

⁷⁴ O título “*Contributions to the Mathematical Theory of Evolution*” ou “*Mathematical Contributions...*” foi empregado como título geral de dezessete relatórios – numerados de II a

“leis” de Mendel foram, entretanto, mencionadas nestes artigos sendo objeto de críticas.⁷⁵ Esse assunto será abordado no próximo capítulo.

Observamos que estes artigos trouxeram contribuições importantes para o desenvolvimento da teoria da análise de regressão ⁷⁶ e do coeficiente de correlação ⁷⁷.

Pearson propôs os fundamentos do sistema de curvas, no segundo relatório da série “Skew Variation in Homogenous Material” (Variação assimétrica em material homogêneo), de 1895. Este sistema inclui curvas simétricas e assimétricas de intervalos limitados e ilimitados, em sua maior parte unimodal.⁷⁸ O propósito de Pearson era fornecer um conjunto de curvas de frequência de diversos formatos⁷⁹ que poderiam ajustar-se aos dados como construções matemáticas para descrever de modo preciso aquilo que observamos.⁸⁰

Pearson escreveu seu primeiro artigo fundamental sobre a teoria da correlação e suas aplicações a problemas de hereditariedade em 1896. O artigo foi intitulado “Regression, Heredity e Panmixia”. Neste trabalho, desenvolveu quase por completo a teoria da correlação normal multivariada como uma ferramenta prática,⁸¹ que aplicou mais tarde, em seu estudo sobre a homotipose. Em 1898, Karl Pearson foi agraciado pela *Royal Society* com a medalha Darwin.⁸²

Em seus estudos, Pearson não procurou investigar distribuições amostrais em pequenas amostras de distribuição de uma população, porque não via nenhuma necessidade para tal. Tanto ele como sua equipe, nas décadas de 1890 e 1900, tomaram por missão o progresso do conhecimento e

XIX. A palavra “mathematical” tornou-se a primeira do título a partir do número III (1896) e assim permaneceu nos seguintes.

⁷⁵ A.W.F. Edwards, “R.A. Fisher on Karl Pearson”, *Notes and Records of the Royal Society of London* 48 (1, 1994):97-106, na p. 102.

⁷⁶ Correlação referente ao desvio das variáveis.

⁷⁷ O quanto uma variável depende da outra.

⁷⁸ Moda é o valor amostral que tem a maior frequência, ou seja, ele é encontrado em maior número de vezes. Uma distribuição que tem apenas uma moda é chamada unimodal.

⁷⁹ O ajuste de curvas de Pearson a dados observacionais foi posto amplamente em prática, nas décadas que se seguiram, por biólogos e cientistas sociais. Os resultados ajudaram muito a dissipar a aceitação da distribuição normal como modelo matemático para descrever a variação de fenômenos biológicos, físicos e sociais.

⁸⁰ Eisenhart, “Pearson, Karl”, na p.451.

⁸¹ *Ibid*, p.452.

⁸² Nos anos pares, a *Royal Society* de Londres concede uma medalha Darwin a pesquisadores que tenham se destacado na área de biologia. Este prêmio foi instituído em 1890.

da compreensão da variação, hereditariedade e seleção em animais e plantas por meio de estudos baseados no exame de números estatisticamente grandes de espécimes e o desenvolvimento da teoria estatística.⁸³ Eles acreditavam que amostragens pequenas não proporcionavam resultados coerentes. O motivo de não despertar interesse a análise de resultados de experimentos em pequena escala poderia ser o fato de que pela preocupação das pequenas diferenças que pretendiam obter, fossem necessárias grandes amostras; ou segundo o filho de Pearson, Egon S. Pearson, representaria uma tentativa de “desencorajar, no biólogo ou médico, a crença de que ele havia recebido um método fácil para tirar conclusões a partir de dados escassos.”⁸⁴ A nosso ver, ambas as possibilidades se aplicam.

Esses métodos e cálculos trouxeram muitos questionamentos por parte dos biólogos da época, por isso no dia 16 de novembro de 1900, Weldon escreveu a Pearson sobre a intenção de fundar um novo periódico. Isso transparece em carta, que foi posteriormente publicada em *Biometrika* na ocasião do falecimento de Weldon.

A afirmação ‘de que números não significam nada e não existem na natureza’ é uma coisa muito séria, contra a qual será necessário lutar. A maior parte das outras pessoas já a superou, mas muitos biólogos ainda não. Você acredita que seria caro demais fundar um periódico de algum tipo...⁸⁵

Em 29 de novembro de 1900, Weldon escreveu a Pearson: “Encontre um nome melhor para esse futuro periódico do que o que consigo imaginar.”⁸⁶, e Pearson respondeu com a sugestão de que “a ciência do futuro deve ser chamada biometria, e seu órgão oficial, *Biometrika*.”⁸⁷

Por meio de um pedido de apoio financeiro que circulou durante dezembro de 1900, foram levantados fundos suficientes para manter o

⁸³ Churchill Eisenhart, “Pearson, Karl”. In: Charles Coulston Gillispie, ed. *Dictionary of scientific biography*. Vol. 9 (New York: Charles Scribner’s sons, 1981): 447 - 473, na p.454.

⁸⁴ Egon S. Pearson, “Some reflexions on continuity in the development of mathematical statistics, 1885-1920, *Biometrika* 54(1967): 341-355, nas pp. 351 - 352.

⁸⁵ Karl Pearson, “Walter Frank Raphael Weldon. 1860–1906”. *Biometrika* 5 (1, 1906): 1–52, na p. 35.

⁸⁶ *Ibid.*

⁸⁷ *Ibid.*

periódico *Biometrika* (1901-1936) por alguns anos. Weldon, Pearson e Charles Benedict Davenport (1886-1944) seriam os editores e Galton atuou como consultor editorial.⁸⁸ Pode-se dizer que, embora Galton tenha aceitado o convite, evitou tomar partido durante a controvérsia entre mendelianos e biometricistas e demonstrou tristeza na ruptura da amizade entre Bateson e Weldon. No primeiro volume de *Biometrika*, publicado em outubro de 1901, intitulado “The Scope of *Biometrika*”, apareceu a proposta desta revista:

- a) Relatórios sobre variação, herança e seleção em animais e plantas, baseados no exame de números estatisticamente grandes de espécimes;
- b) Desenvolvimentos da teoria estatística que sejam aplicáveis e problemas biológicos;
- c) Tabelas numéricas e soluções gráficas que levem a reduzir o trabalho de aritmética estatística;⁸⁹
- d) Sumários de relatórios que tratem desses temas e estejam publicados em outros locais;
- e) Notas sobre trabalhos atuais em biometria e problemas ainda sem solução.

Nos anos que se seguiram, a *Biometrika* tornou-se um dos veículos para a publicação de tabelas matemáticas e outros recursos de auxílio à análise estatística, bem como tabelas detalhadas de dados biológicos.⁹⁰

A morte prematura de Weldon em 1906 ocasionou uma mudança significativa na vida de Pearson. Além do grande choque, teve a perda de um grande amigo e colaborador, sem o qual sentiu dificuldades no prosseguimento do trabalho na área da biometria. Porém, Pearson continuou produzindo em seus trinta anos restantes, mais de trezentos artigos, editoriais, relatórios e livros sobre biometria e estatística, ou sobre temas relacionados. Pearson também escreveu um livro sobre astronomia e quatro sobre mecânica, além de cerca de setenta cartas, resenhas e notas em publicações científicas; a última

⁸⁸ Churchill Eisenhart, “Pearson, Karl”. In: Charles Coulston Gillispie, ed. *Dictionary of scientific biography*. Vol. 9 (New York: Charles Scribner’s sons, 1981): 447 - 473, na p. 463.

⁸⁹ Atualmente, nos livros de estatística, é feita referência das tabelas numéricas de Pearson.

⁹⁰ Eisenhart, “Pearson, Karl”, na p. 463.

carta (1935) foi a respeito dos objetivos dos fundadores da *Biometrika* e das condições sob as quais o periódico havia sido publicado.⁹¹

Em 1911, Pearson foi o primeiro a assumir a cátedra do “National Eugenics” de Francis Galton no *University College* de Londres, onde permaneceu até a sua aposentadoria como professor emérito até 1933⁹². Em julho de 1932, Pearson informou a universidade de que abriria mão da cátedra Galton no verão seguinte. A faculdade decidiu dividir o departamento de matemática aplicada em duas unidades independentes: um departamento de eugenia, com o qual a cátedra Galton estaria associada, e um novo departamento de estatística. Em outubro de 1933, o departamento de zoologia pôs uma sala à disposição de Pearson e de seu filho, Egon Pearson, que tornou-se chefe desse novo departamento de estatística; R.A.Fisher foi nomeado o segundo professor Galton de eugenia. Pearson continuou a editar a *Biometrika* e chegou a levar ao prelo a primeira metade das provas finais daquele que seria o volume 28, antes de falecer aos 79 anos de idade, em 27 de abril de 1936.⁹³

⁹¹ Churchill Eisenhart, “Pearson, Karl”. In: Charles Coulston Gillispie, ed. *Dictionary of scientific biography*. Vol. 9 (New York: Charles Scribner’s sons, 1981): 447 - 473, na p. 463.

⁹² A.W.F. Edwards, “R.A. Fisher on Karl Pearson”, *Notes and Records of the Royal Society of London* 48 (1, 1994): 97-106, na p. 101.

⁹³ Churchill Eisenhart, “Pearson, Karl”. In: Charles Coulston Gillispie, ed. *Dictionary of scientific biography*. Vol. 9 (New York: Charles Scribner’s sons, 1981): 447 - 473, na p. 464.



Figura 2.1: Karl Pearson

(Fonte: A.W.F. Edwards. *R.A.Fisher on Karl Pearson*. Notes and Records of the Royal Society of London. Vol. 48, nº1 (Jan. 1994), p.98)

Abaixo, segue uma tabela com informações que aparecem de modo esparso em Churchill Eisenhart.

KARL PEARSON: CRONOLOGIA

ANOS	IDADE	
1879	22	Recebeu o título de bacharel com honras matemáticas pela Cambridge University.
1880	23	Publicação: "The New Werther" .
1881	24	Recebeu o título de bacharel em Direito pela Universidade de Cambridge.
1884	27	Foi indicado para a cátedra Goldsmid de matemática aplicada e mecânica no <i>University College</i> de Londres
1886	29	Publicação: "The Ethic of Freethought"
1887	30	Publicação: "Die Fronica"
1890	33	Casou-se com Maria Sharpe, com quem teve um filho, Egon, e duas filhas, Sigrid e Helga.
1891-1894	34-37	Lecionou em cursos no <i>Gresham College</i>
1892	35	Publicação: <i>The Grammar of Science</i> .
1893	36	Em 31 de janeiro o termo desvio-padrão foi introduzido em uma aula. Publicações: "The Geometry of Chance" e "Asymmetrical frequency curve" Ministrou aulas sobre variação e sobre correlação para alunos pesquisadores do University College.
1893-1912		Escreveu um conjunto de 18 artigos denominado "Mathematical Contribution to the Theory Evolution", com contribuições importantes para o desenvolvimento da teoria da Análise de Regressão e do Coeficiente de Correlação.
1894	37	Publicação: "On the dissection of asymmetrical frequency curves."
1895	38	Publicação: "Skew variation in homogeneous material."
1896	39	Foi eleito membro da <i>Royal Society of London</i> . Publicação: "Regression, heredity and panmixia."
1897	40	Publicação "The chances of death and other studies in evolution"
1898	41	Recebeu a medalha Darwin.
1899-1900	42-43	Coletou material para testar uma nova teoria da homotipose.
1901	44	Fundou o jornal <i>Biometrika</i> com Weldon e Francis Galton.
1906	49	Publicação: "A mathematical theory of random migration"
1909	52	Publicação: "The problem of practical eugenics e Tresury of Human Inheritance", no qual promoveu um valioso material de anomalias raras.
1910	53	Publicações: "A preliminary study of extreme alcoholism in adults" e "The influence of parental alcoholism on the physique and ability of the offspring".
1911	54	Abriu mão da cátedra <i>Goldsmid</i> de matemática aplicada para tornar-se o primeiro a assumir a cátedra Galton de eugenia e

		chefe do novo departamento de estatística aplicada. Publicação: “A monograph on albinism in man”.
1912	55	Publicação: “Darwinism, medical progress and eugenics: The Cavendish lecture, an address to the medical profession”
1913	56	Publicação: “On the correlation of fertility with social value: a cooperative study”
1914	57	Publicações: “Tables for statisticians and biometricians, Mendelism and the problem of mental defect” e “The life, letters and labours of Francis Galton.”
1920	63	Publicação: “On the construction of tables and on interpolation.”
1922	65	Publicação: <i>Francis Galton, 1822-1922.</i>
1923	66	Publicação: <i>Charles Darwin, 1809-1882.</i>
1925	68	Fundou a revista <i>Annals of Eugenics.</i>
1928	71	Morte de sua esposa Maria Sharpe.
1929	72	Casou-se com Margaret Victoria Child.
1933	76	Aposentou-se, mas continuou a trabalhar em uma sala do <i>University College.</i>
1934	77	Publicação: “Tables of the incomplete beta-function”

2.2 MENDELIANOS VERSUS BIOMETRICISTAS

Como já mencionamos anteriormente, após a morte de Darwin (1882), vários cientistas voltavam seus estudos para os esclarecimentos sobre a seleção natural. Martins comentou a respeito:

...estudiosos de diferentes países estavam preocupados com a questão da origem das espécies. Eles queriam obter esclarecimentos acerca de como ocorriam as variações, qual o tipo de variação que seria mais relevante para o processo evolutivo e de que modo elas seriam preservadas. Além disso, desejavam testar a eficácia da seleção natural.⁹⁴

Seguindo uma das sugestões de Darwin sobre os estudos que poderiam ser feitos para o esclarecimento da origem das espécies, vários cientistas

⁹⁴ Lilian Al-Chueyr Pereira Martins, “Weldon, Pearson, Bateson e a controvérsia mendeliano-biometricista: uma disputa entre evolucionista”, *Filosofia Unisinos* 8 (2, 2007): 170-190, na p. 171.

procuraram reconstruir a filogenia, através dos estudos embriológicos. Nesse sentido, destacou-se a escola de Cambridge liderada por Frank Balfour. Dela faziam parte William Bateson (1861-1926) e Walter Frank Raphael Weldon, que eram amigos muito próximos. Entretanto, com a morte prematura de Balfour em um acidente de alpinismo e a insatisfação com os resultados das investigações dentro dessa linha, a escola de Cambridge se dispersou. Bateson se dedicou a outras linhas de investigação, voltando-se, no final da década de 1890, para o estudo de cruzamentos experimentais de animais e plantas⁹⁵.

Os trabalhos de Weldon, inicialmente, estavam relacionados à fauna marinha e somente depois ele adotou um tratamento estatístico em suas pesquisas⁹⁶. Durante a década de 1890, Weldon interagiu de maneira intensa com Karl Pearson, desenvolvendo com ele uma série de estudos biológicos onde aplicava métodos estatísticos, iniciando com suas investigações sobre a variação no camarão *Crangon vulgaris*, onde calculou os primeiros coeficientes de correlação orgânica. Pearson e Weldon publicaram vários trabalhos onde aplicavam os métodos biométricos e contavam com diversos colaboradores e alunos que trabalhavam dentro desta linha. Mesmo quando Weldon se mudou de Londres para Oxford (1899), eles mantiveram sua parceria. Normalmente, em seus estudos, Weldon e Pearson lidavam com características que eram herdadas de forma contínua e, nesse caso, o papel da seleção natural era fundamental⁹⁷. Neste cenário, houve um acontecimento importante: a “redescoberta” do trabalho de Mendel em 1900 por Hugo de Vries, Carl Correns e Erick Von Tschermak. Bateson trabalhava com cruzamentos experimentais de animais e plantas desde 1897.

Segundo Provine, foi em 8 de maio de 1900, quando Bateson estava a caminho da *Royal Horticultural Society* para fazer uma apresentação sobre os problemas da hereditariedade relacionados à investigação em horticultura. Na viagem de trem de Cambridge para Londres tomou contato com o artigo de

⁹⁵ Lilian Al-Chueyr Pereira Martins, “William Bateson: da evolução à genética”, *Episteme* (8, 1999): 67-88.

⁹⁶ Martins. “Weldon, Pearson, Bateson e a controvérsia mendeliano-biometricista: uma disputa entre evolucionista”, na p. 172.

⁹⁷ *Ibid*, na p. 182.

Mendel sobre ervilhas.⁹⁸ Bateson, que já trabalhava na mesma linha que Mendel, antes mesmo de conhecer seus estudos, dedicou-se a partir daí ao programa de pesquisa mendeliano, onde procurava testar se os mesmos princípios que haviam sido descritos por Mendel em seu artigo sobre ervilhas se aplicavam a outras plantas ou animais. Além disso, ele procurou investigar a existência de desvios ou mesmo propor novas leis⁹⁹. Ocorre que, nos experimentos desenvolvidos por Bateson e outros investigadores que seguiam essa linha, eram estudadas características principalmente descontínuas¹⁰⁰.

Nesse caso, o papel da seleção natural ficava enfraquecido. Bateson começou a atrair estudantes para trabalharem com ele e a publicar seus resultados. Bateson publicou um artigo que foi criticado por Weldon. Em seu artigo, Weldon fazia restrições não só às ideias de Mendel, mas também à interpretação e metodologia de Bateson. Bateson respondeu às críticas do colega com a publicação do livro *Mendel's principles of heredity, a defence*. A partir daí as divergências metodológicas, conceituais (em relação ao tipo de variação que seria mais relevante para o processo evolutivo, à concepção de herança, por exemplo) de terminologia, bem como outros fatores contribuíram para acirrar a controvérsia que ocorreu principalmente entre 1902 e 1906, envolvendo por um lado William Bateson e de outro Weldon e Pearson¹⁰¹. O conflito entre os mendelianos e os biometricistas (ancestralistas) que se deu em publicações, encontros científicos e correspondência é conhecido como controvérsia mendeliano-biometricista e já foi objeto de estudo de vários historiadores e filósofos da ciência.¹⁰²

As críticas vinham de ambas as partes. Segundo Martins, Pearson também questionou as concepções biológicas de Bateson, acusando-o de usar livremente termos como variação, descontinuidade e diferenciação, dizendo

⁹⁸ Mendel utilizou, em suas experiências com ervilhas, somente características descontínuas, mas também descreveu uma experiência com duas variedades de *Phaseolus*. Uma variedade tinha flores brancas e a outra, roxa. Quando cruzadas, todos os híbridos produziram flores roxas. As sementes dos híbridos produziram plantas com flores de uma série de cores, do vermelho púrpura, ao lilás e ao branco. Mendel concluiu que, mesmo esses resultados enigmáticos, porém, poderiam ser provavelmente explicados pela lei governante Pinsum, se pudesse assumir que as cores das flores e as sementes do Ph, são a combinação de duas ou mais cores, inteiramente independentes, as quais agem individualmente, como qualquer outra característica constante na planta.

⁹⁹ Lilian Al-Chueyr Pereira Martins, "Bateson e o programa de pesquisa mendeliano", *Episteme*.

¹⁰⁰ William B. Provine, *The origins of theoretical population Genetics*, p.58.

¹⁰¹ *Ibid*, na p. 57.

¹⁰² *Ibid*.

que o emprego de tais palavras era apresentado sem definição e em sentidos diferentes. E que por sua vez, Bateson não havia utilizado os termos no sentido biométrico porque não tinha um treino em biometria¹⁰³, como Pearson esperava que o tivesse.

Pearson questionou, em relação às características obtidas nos experimentos, que a mudança descontínua poderia surgir de uma feliz combinação de variações contínuas de várias características, as quais ele descrevia geometricamente como um deslocamento na superfície de várias dimensões.¹⁰⁴ Segundo Churchill, os pesquisadores ligados à biometria afirmavam que a evolução de novas espécies resultava de um acúmulo gradual de efeitos de pequenas variações contínuas¹⁰⁵, de modo análogo a Darwin.

A controvérsia entre os mendelianos e biometricistas se estendeu por anos, tanto em aspectos referentes à hereditariedade quanto à evolução. Mas, além disso, houve uma luta pela autoridade no campo da hereditariedade e evolução entre biometricistas e biólogos¹⁰⁶. Isso fica evidente na primeira edição da revista *Biometrika*¹⁰⁷ em 1901:

Há poucos anos, todos aqueles problemas cujas soluções dependiam de um estudo de diferenças entre membros individuais de uma raça ou espécies eram negligenciados por muitos biólogos...

E mais no segundo artigo:

... o problema da evolução é um problema de estatística... devemos voltar à matemática de grandes números, à teoria do fenômeno de massa, a interpretar cuidadosamente nossas observações... não devemos questionar como elas surgiram e nem como o fundador da teoria moderna de descendentes contribuiu tão pouco para a estatística?... As características do pensamento de Darwin o levaram

¹⁰³ Martins, "Karl Pearson, William Bateson e a controvérsia da Homotipose", p. 6.

¹⁰⁴ Theodore M. Porter. *Karl Pearson: the scientific life in a statistical age*. New Jersey: Princeton, 1953, 258.

¹⁰⁵ Eisenhart, "Pearson, Karl", na p.463.

¹⁰⁶ Martins "Weldon, Pearson e Bateson e a controvérsia mendeliano-biometricista: uma disputa entre evolucionistas", p. 188.

¹⁰⁷ Karl Pearson, "The Scope of Biometrika", *Biometrika* 1 (1, 1901-1902) pp. 1-2.

a estabelecer a teoria da descendência comum sem levar em conta as concepções matemáticas...

Além de todos esses aspectos que contribuíram para o prolongamento da controvérsia entre mendelianos e os biometricistas, vemos que “enquanto Bateson procurava valorizar a contribuição de Mendel..., os biometricistas procuravam valorizar a contribuição de Galton e o método biometricista”¹⁰⁸. Em meio a essa discussão, existia a falta de empenho de ambos os lados em procurar entender os conceitos e terminologia empregados pela outra parte.¹⁰⁹

Segundo Edwards, a controvérsia só mostrou que Bateson não sabia o suficiente de matemática e Pearson não sabia sobre biologia, e que o campo da investigação estatística aberta pela teoria de Mendel não foi reconhecido por ambos.¹¹⁰

Alguns elementos que fizeram parte da controvérsia entre mendelianos e biometricistas já apareceram antes da controvérsia propriamente dita ter se iniciado, como veremos a seguir na discussão entre Pearson e Bateson sobre o princípio da homotipose. Pearson iniciou seus estudos referentes aos órgãos similares indiferenciados em 1899. Na ocasião, coletou um grande volume de materiais do reino vegetal para verificação de sua hipótese sobre a homotipose. O artigo de Pearson sobre o assunto gerou uma discussão entre ele e Bateson, o que acabou gerando um conflito maior, embora, o próprio Galton, inspirador da escola, não tenha se envolvido diretamente na discussão a respeito da natureza da evolução.¹¹¹

Mesmo havendo muitas divergências¹¹² sobre a homotipose, Bateson concordou com alguns dentre os pressupostos da homotipose¹¹³, tais como:

¹⁰⁸ Lilian Al-Chueyr Pereira Martins. “Weldon, Pearson, Bateson e a controvérsia mendeliano-biometricista: uma disputa entre evolucionista”, *Filosofia Unisinos* 8 (2,2007): 170-190, na p. 187.

¹⁰⁹ *Ibid*, na p.188.

¹¹⁰ A.W.F. Edwards, “R.A. Fisher on Karl Pearson”, *Notes and Records of the Royal Society of London* 48 (1, 1994): 97-106, na p. 103.

¹¹¹ Eisenhart, “Pearson, Karl”, na p. 463.

¹¹² A teoria de Mendel foi contraditória à teoria da homotipose de Pearson. Para Mendel os descendentes dos híbridos eram diferenciados porque as células dos embriões eram diferenciadas. Uma célula do embrião era diferente da outra, porque tinha uma combinação diferente de elementos diferenciados.

¹¹³ Bateson não usou a teoria de Mendel em suas críticas, pois não acreditava que os elementos diferenciados de Mendel fossem corpos materiais, desta forma ele rejeitou a teoria da hereditariedade dos cromossomos.

... o relacionamento e a semelhança entre dois irmãos são a expressão de um mesmo fenômeno, como o relacionamento e a semelhança entre duas folhas da mesma árvore, entre as escamas das asas de uma mariposa, as pétalas de uma flor, e, geralmente, entre partes repetidas.¹¹⁴

Apesar de concordar com esse aspecto e valorizar “a contribuição feita pelos biometricistas nos últimos anos”¹¹⁵, Bateson fez críticas sobre a ausência de distinção teórica de Pearson, entre diferenciação e variação num único indivíduo ou população e a ignorância da importância das variações contínuas e descontínuas neste estudo.¹¹⁶

Embora não tenha envolvido diretamente a hereditariedade mendeliana, contribuiu para o início da luta entre os biometricistas e os mendelianos.

Como vimos anteriormente, após o episódio relacionado à publicação do artigo sobre a homotipose, Pearson e Weldon não concordaram com os procedimentos adotados pela *Royal Society* e decidiram criar uma nova revista científica, a *Biometrika*.¹¹⁷ Na próxima seção discutiremos sobre os conceitos relacionados ao princípio da homotipose de Karl Pearson.

2.3 CONCEITOS BIOLÓGICOS DE KARL PEARSON E O PRINCÍPIO DA HOMOTIPOSE NO INÍCIO DO SÉCULO XX

Como mencionamos anteriormente, Karl Pearson trabalhou em sua teoria da homotipose no verão de 1899. Coletou material e, em 6 de outubro de 1900, submeteu um resumo à *Royal Society* e o leu em 15 de novembro do

¹¹⁴ William Bateson, “Heredity, Differentiation, and Other Conceptions of Biology: A consideration of Professor Karl Pearson’s Paper ‘On the Principle of Homotyposis’”, *Royal Society* 69 (1901): 1 – 404.

¹¹⁵ Lilian Al-Chueyr Pereira Martins, “Weldon, Pearson, Bateson e a controvérsia mendeliano-biometricista: uma disputa entre evolucionista”, *Filosofia Unisinos* 8 (2,2007): 170-190, na p. 175.

¹¹⁶ Provine, *The origins of theoretical population Genetics*, p. 62.

¹¹⁷ *Ibid.*

mesmo ano¹¹⁸. Depois publicou o artigo intitulado “Mathematical contributions to the theory of evolution IX”. De acordo com Provine, o artigo “foi uma tentativa de Pearson para simplificar o problema da hereditariedade.”¹¹⁹

Para este trabalho, Pearson coletou um maior volume de dados em comparação aos seus estudos anteriores.

Pearson observou a existência de órgãos similares indiferenciados, como as folhas de uma mesma árvore, e sentiu necessidade em introduzir novas terminologias para descrever suas ideias sobre grau de semelhança entre órgãos similares indiferenciados. Em suas palavras:

... tanto estou consciente, [de que se trata] de um novo campo de investigação, que me aventuro com alguma hesitação a introduzir certos termos para descrever ideias frequentemente decorrentes. Devo chamar os órgãos similares indiferenciados de homotipos, indicando que são produzidos pelo mesmo molde ou indivíduo. Assim duas folhas da mesma árvore, ou duas células sanguíneas do mesmo sapo são homotipos, suas semelhanças serão homotípicas.¹²⁰

Assim, para Pearson, esse grau de semelhança entre os órgãos similares indiferenciados foi chamado de homotipose.

Em relação à variabilidade e correlação dos órgãos similares indiferenciados, em seu artigo, Pearson adotou os métodos de seleção e recolhimento do material de pesquisa, os processos estatísticos empregados na redução do material e a terminologia adotada (homotipose e homotipos). Optou pela coleta de materiais que fossem fáceis de contar e medir, e explicitou quais seriam as características que iria examinar em seus órgãos. Nesse sentido, ele esclareceu:

Examino meus indivíduos e seus órgãos por sinais externos óbvios para informação, mas não estudo a evolução morfológica dos órgãos

¹¹⁸ Provine. *The origins of theoretical population Genetics*, p. 58.

¹¹⁹ *Ibid*, p.59.

¹²⁰ Pearson, “Mathematical Contributions to the Theory of Evolution IX”, na p. 294.

considerados, ou questiono se as partes contadas foram todas originadas de uma mesma fonte.¹²¹

Pearson contou o número de veias das folhas como ilustrado em seu outro artigo “On differentiation and homotyposis in the leaves of *Fagus sylvatica*”¹²²

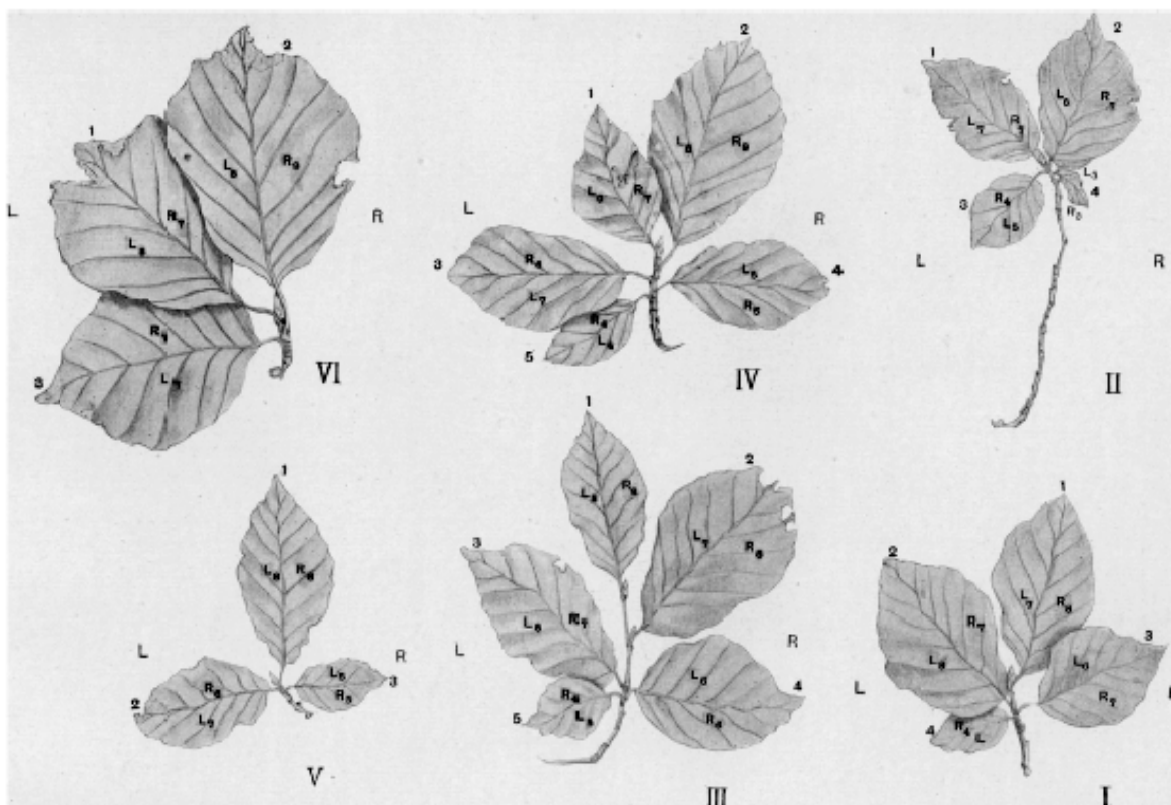


Figura 2.1 – Folhas da Shade Branch

(Fonte: Karl Pearson, Marion Radford, “On differentiation and homotyposis in the leaves of *Fagus sylvatica*”, *Biometrika* 3 (1, 1904):104-107, na p.106)

O resultado deste estudo sobre homotipose foi um artigo em coautoria com diversos membros de sua equipe¹²³, “On the Principle of Homotyposis and Its Relation to Heredity... Part I. Homotyposis in the Vegetable Kingdom”. Eles coletaram material suficiente para elaborar “22 séries com 29 tabelas de correlação”¹²⁴. Pearson e sua equipe recolheram folhas da faia, em uma

¹²¹ Pearson, “Mathematical Contributions to the Theory of Evolution IX”, na p. 291.

¹²² Karl Pearson & Marion Radford, “On Differentiation and Homotyposis in the Leaves of *Fagus sylvatica*”, *Biometrika* 3 (1, 1904): 104-107, na p.106.

¹²³ Sua equipe era formada por Alice Lee, C.D. Fawcett, Leslie Bramley-Moore e outros.

¹²⁴ Martins, “Karl Pearson, William Bateson e a controvérsia da Homotipose”, p. 3.

centena de árvores de mesma idade e reuniram de cada árvore 26 folhas. Por fazerem a coleta verificando características como a região, mesma altura e diferentes partes do ramo da árvore, supunham que cada árvore era individualizada pelas 26 folhas. As veias nessas folhas foram contadas e variadas entre 10 e 22. Calcularam todos os pares possíveis $\frac{1}{2} (26 \times 25) = 325$ em número; como a tabela de correlação foi representada simetricamente considerando pelo início de qualquer uma das folhas no par como primeiro ou segundo, então, uma árvore levou 650 entradas. O total de 100 árvores estudadas resultou em 650.000 entradas¹²⁵. Neste estudo, Pearson e sua equipe coletaram, contaram e calcularam um grande número de amostras. Pearson coletou outras folhas além da faia, como as da papoula e do azevinho, todas foram calculadas e tabuladas.

Nesses estudos, além dos cálculos e tabelas estatísticas, Pearson descreveu detalhadamente os órgãos indiferenciados semelhantes no indivíduo, o termo homotipose e sua relação com a hereditariedade. Um dos aspectos analisados por Karl Pearson neste trabalho foi a variabilidade em relação à seleção e as características dos descendentes em relação a um único pai:

... tenho frequentemente ficado impressionado com a pequena redução na variabilidade produzida pela seleção. Os descendentes de um único pai enquanto divergem em características, ainda terão uma variabilidade somente um pouco reduzida, no máximo 10% abaixo da variabilidade racial. (...) Agora essa capacidade do pai em produzir descendentes variáveis deve estar de alguma maneira relacionada ao grau de semelhança do descendente.¹²⁶

Karl Pearson testou uma variedade de dados deste relacionamento teórico, e sua discussão estava voltada na suposição de que a correlação fraternal igualava-se à correlação homotípica, isto porque tais dados resultaram em um valor principal de correlação homotípica de 0,4570 e Pearson já havia

¹²⁵ Pearson, "Mathematical Contributions to the Theory of Evolution IX", na p. 293.

¹²⁶ *Ibid*, na p. 286.

obtido um valor de 0,4479 para correlação fraternal. Assim, correlação homotípica e fraternal eram, para ele, “sensivelmente iguais”¹²⁷.

Neste estudo, Pearson assumiu que, mesmo os órgãos semelhantes, considerados indiferenciados, ainda apresentam diferenças entre si. Seu conceito sobre a variabilidade dos órgãos similares indiferenciados se baseia na combinação de pequenas razões hereditárias e ambientais.¹²⁸ Nas palavras de Pearson:

Não afirmamos que muito ou pouco seja devido ao ambiente e que muito ou pouco é devido à influência da hereditariedade ancestral. Tudo que assumimos é que isso causa a semelhança entre as folhas da mesma árvore ou brotos da mesma flor, produzem a semelhança entre espermatozóide e óvulo de um mesmo indivíduo e a semelhança de um descendente do mesmo pai.¹²⁹

A partir desta posição, Pearson investigou o quanto estas semelhanças dos descendentes de um mesmo pai estão relacionadas à semelhança dos órgãos similares não diferenciados gerados por um indivíduo.

Pearson discutia que um organismo individual produz “órgãos similares indiferenciados”, tais como corpúsculos sanguíneos, pétalas de flores ou flores de árvores, e que esses órgãos não são exatamente semelhantes, porém possuem certa variabilidade que é menor que a variabilidade dos órgãos similares em uma raça. Para Pearson, a variação dentro do indivíduo é menor que a da raça, ou que os órgãos semelhantes indiferenciados têm certo grau de semelhança.¹³⁰ E chamou de Homotipose “o princípio que os homotipos

¹²⁷ Provine. *The origins of theoretical population Genetics*, p.60.

¹²⁸ Karl Pearson, “Mathematical Contributions to the Theory of Evolution. IX. On the Principle of Homotyposis and Its Relation to Heredity, to the Variability of the Individual, and to that of the Race. Part I. Homotypos in the Vegetable Kingdom”, *Philosophical Transactions of the Royal Society of London* 197 (1901): 285-379, na p. 287.

¹²⁹ *Ibid*, na p. 288.

¹³⁰ Neste trabalho, Pearson discutiu as questões, se: (1) um único organismo produz órgãos similares indiferenciados, (2) as células do espermatozóide e do óvulo, podem ser razoavelmente considerados como órgãos similares indiferenciados, (3) os descendentes são representantes fiéis dos germens parentais, (4) a semelhança quantitativa entre o descendente de um mesmo pai deve ser o mesmo quanto à semelhança quantitativa entre órgãos similares indiferenciados num organismo individual.

são correlacionados”¹³¹, que seria o grau de semelhança entre os órgãos similares indiferenciados. Para Pearson, a “variação dentro de um indivíduo é menor do que a da raça”¹³².

Neste estudo sobre homotipose, Pearson relatou que a concepção teórica dos órgãos similares indiferenciados era muito difícil de se concretizar na prática e considerou alguns fatores que poderiam prejudicar sua amostragem, como recolhimento do material e o fator ambiental, dizendo:

... a posição do órgão sobre o ramo, ou do ramo de uma planta, mesmo quão cuidadoso o coletor poderá ser, poderia realmente ter introduzido a diferenciação e assim enfraquecido a homotipose aparente. Em segundo lugar, o fator ambiental aparece. É difícil obter uma centena de indivíduos no mesmo ambiente, solo, (...), luz solar,...) ou ao contrário, o ambiente pode produzir uma semelhança fictícia nos órgãos do mesmo indivíduo, quando passamos de um indivíduo para outro. O fato de que metade de nossas séries cresceu num campo, e outra metade em outro, e que parte veio de um lado da estrada, e uma parte de outra, isso pode introduzir uma heterogeneidade despercebida. Em terceiro lugar, a dificuldade em garantir que todos os indivíduos tenham a mesma idade, ou o mesmo estágio de desenvolvimento, é muito grande. (...) Estes e outros fatores talvez possam ser eliminados até certo ponto – muito melhor por um botânico treinado do que por um simples matemático – mas eu duvido muito sobre a possibilidade de sua eliminação completa.¹³³

Pearson verificou se, mesmo com tais fatores, a homotipose tinha um valor praticamente constante em toda natureza, e deveria apenas considerar como resultado um valor médio de todas as séries, em que tais fatores de diferenciação se opõem (ambiente, idade, fase de crescimento) podendo neutralizar cada um.

¹³¹ Pearson, “Mathematical Contributions to the Theory of Evolution. IX”, na p. 294.

¹³² *Ibid.*

¹³³ *Ibid*, pp. 358 - 359.

No capítulo 3 trataremos dos aspectos relacionados à análise estatística propriamente dita.

CAPÍTULO 3

ANÁLISE MATEMÁTICA DO TRABALHO DE KARL PEARSON SOBRE HOMOTIPOSE

Este capítulo analisa alguns aspectos estatísticos relacionados ao princípio da homotipose. Considera a coleta de materiais, tabulações e o método de verificação de resultados presentes no artigo “Mathematical Contributions to the Theory of Evolution. IX.” (1901). A escolha deste artigo foi por se tratar do primeiro trabalho de Pearson sobre homotipose, no qual ele e sua equipe apresentaram uma vasta pesquisa e coleta de dados do reino vegetal.

3.1 CONCEITOS ESTATÍSTICOS E A CORRELAÇÃO

Desde seus estudos em *Grammar of Science*, Pearson aceitava a generalização e a forma de medir a precisão.¹³⁴ Segundo o autor, “a missão da ciência não é explicar, mas descrever e descobrir a fórmula que capacitará o homem a prever a natureza das percepções futuras.”¹³⁵ Nesse sentido, a estatística se distingue das associações aleatórias e deve lidar essencialmente com os fenômenos.¹³⁶

Antes de se dedicar a homotipose, Pearson interessou-se pela hereditariedade e evolução, o que transparece em seu primeiro artigo sobre

¹³⁴ Theodore M. Porter. *Karl Pearson: the scientific life in a statistical age*. Princeton University Press, New Jersey: 1953, p.257.

¹³⁵ Pearson, “Reaction,” in *Chances of Death*, p.200.

¹³⁶ Theodore M. Porter. *Karl Pearson: the scientific life in a statistical age*. Princeton University Press, New Jersey: 1953, p.257.

correlação e regressão.¹³⁷ Neste, para ilustrar o valor dessas novas ferramentas matemáticas na abordagem de problemas nessa área, utilizou estimativas do trabalho de Galton, “Record of Family Faculties”¹³⁸ acerca das regressões parciais dos descendentes sobre cada progenitor.¹³⁹ . Pearson e Weldon aderiram à “lei da herança ancestral” de Galton, formulada em *Natural Inheritance* (1889). Esta lei, segundo Churchill, era de certo modo ambígua e imprecisa, por não levar em consideração a complexidade matemática adicional envolvida na abordagem conjunta de mais de duas características mutuamente correlacionadas.¹⁴⁰

Um tópico importante, sobre a teoria quantitativa da hereditariedade baseada na estatística foi a fundamentação apresentada por Galton (1897) através de sua lei da herança ancestral. Ela aparece sob a forma de uma equação de regressão múltipla dos descendentes sobre a ancestralidade média dos pais. A partir de seus estudos sobre a herança de pelagem em cachorros (*Basset-hounds*), Galton admitia que os pais contribuíam com a média de $\frac{1}{2}$ ou (0,5), os avós com $\frac{1}{4}$ ou (0,5²), os bisavós com $\frac{1}{8}$ ou (0,5³), e assim por diante¹⁴¹ o que é representado pela equação:

$$x_0 = \frac{1}{2} \frac{\sigma_0}{\sigma_1} x_1 + \frac{1}{4} \frac{\sigma_0}{\sigma_2} x_2 + \frac{1}{8} \frac{\sigma_0}{\sigma_3} x_3 + \dots$$

O total da soma representava toda a herança que correspondia a 1.

Onde:

x_0 é o desvio previsto de um descendente individual em relação à média da geração de descendentes;

x_1 é o desvio da média dos pais do descendente em relação à média da geração parental;

¹³⁷ Karl Pearson, “Mathematical contributions to the theory of evolution. III. Regression, heredity and panmixia”, *Philos. Trans. Roy. Soc. London Ser. A* 187 (1896): 253 - 318.

¹³⁸ O “Record of Family Faculties” era um questionário sobre os dados de famílias pesquisadas, mostrando características como altura, habilidades físicas, mentais e artísticas, cor de cabelo, doenças, etc.

¹³⁹ Eisenhart, “Pearson, Karl”, na p. 462.

¹⁴⁰ *Ibid.*

¹⁴¹ Francis Galton, “The average contribution of each several ancestor to the total heritage of the offspring.” *Proceeding of the Royal Society* 61 (1897): 401-413, na p. 402.

x_2 é o desvio da média dos avós do descendente em relação à média da geração dos avós, e assim por diante;

$\sigma_0, \sigma_1, \dots$ são os desvios-padrão das distribuições dos indivíduos em suas respectivas gerações.¹⁴²

Para Churchill era preciso que Galton fornecesse uma definição precisa de “média dos s-ésimos pais”.¹⁴³ Para Pearson, a definição adotada foi se “um pai é o primeiro pai, um avô o segundo pai, e um bisavô o terceiro pai, e assim por diante, então a média dos s-ésimos pais é a média de todos os 2^{os} s-ésimos pais individuais”

A partir deste raciocínio, Pearson atribuiu valores teóricos para os coeficientes de correlação e regressão entre diversas relações de parentescos. Testou os dados de estaturas coletados por Galton com relação a estas estimativas, e sugeriu generalizar a lei de Galton introduzindo algumas variáveis ($\gamma\beta, \gamma\beta^2, \gamma\beta_3, \dots$) no lugar dos coeficientes de série geométrica de Galton ($\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \dots$). Esta foi uma forma que permitiu “uma maior abrangência para a variedade da herança em diferentes espécies”

Segundo Churchill algumas afirmações de Pearson sobre a equação da hereditariedade ancestral, davam uma impressão errada de que a lei expressava uma relação entre um indivíduo particular e seus s-ésimos pais. Isto teria confundido alguns biólogos da época que não estavam totalmente conscientes de que equações de regressão expressavam relações que só se aplicam à média entre tipos genéricos de indivíduos envolvidos, e não entre indivíduos particulares desses tipos.¹⁴⁴

Nos estudos relacionados à variabilidade da população e correlação, Galton também notou que a variação no tamanho dos descendentes dentro de cada vetor parental era aproximadamente constante e menor do que aquela que, medida do mesmo modo, podia ser verificada em toda população de descendentes. Ele fez então a relação:

$$\text{Variabilidade da família} = \sqrt{1 - r^2} \times \text{variabilidade de toda a população}$$

¹⁴² Churchill Eisenhart, “Pearson, Karl”. In: Charles Coulston Gillispie, ed. *Dictionary of scientific biography*. Vol. 9 (New York: Charles Scribner’s sons, 1981): 447 - 473, na p. 462.

¹⁴³ *Ibid.*

¹⁴⁴ *Ibid.*

Nesta fórmula Galton, em 1877, considerou a variável r que chamou de coeficiente de reversão, como dito no capítulo 1. O cientista inglês continuou trabalhando com esta fórmula em sua obra *Natural Inheritance* (1889) como aparece na tabela abaixo. O que Galton inicialmente chamou de reversão, passou mais tarde a chamar de regressão.¹⁴⁵ Aplicou este conceito para calcular a regressão e variabilidade em relação à estatura.

TABLE OF DATA FOR CALCULATING TABLES OF DISTRIBUTION OF STATURE AMONG THE KINSMEN OF PERSONS WHOSE STATURE IS KNOWN.

From group of persons of the same Stature, to their Kinsmen in various near degrees.	Mean regression= w .	$Q = f = p \times \sqrt{1 - w^2}$.
Mid-parents to Sons	2 / 3	1.27
Brothers to Brothers	2 / 3	1.27
Fathers or Sons to } Sons or Fathers }	1 / 3	1.60
Uncles or Nephews to } Nephews or Uncles }	2 / 9	1.66
Grandsons to Grandparents...	1 / 9	Practically that of Population, or 1.7 inch.
Cousins to Cousins	2 / 27	

Figura 3.1 – Tabela referente à distribuição da estatura

(Fonte: Francis Galton, *Natural Inheritance*, p.133)

Pearson utilizou esta fórmula em seu trabalho sobre homotipose¹⁴⁶. Comparou seus resultados de correlação fraternal que haviam sido obtidos a partir dos dados coletados, ao resultado do cálculo desta fórmula. Chegou a valores¹⁴⁷ que tinham concordância com os valores médios da tabela (valores que veremos no final deste capítulo).

Tanto nos trabalhos de Pearson como nos de Galton, encontramos casos que envolviam duas variáveis em uma população. Eles procuravam determinar se havia alguma relação entre essas variáveis. Em caso positivo,

¹⁴⁵ Karl Pearson, "Notes on the History of Correlation", *Biometrika*, 13 (1, 1920): 25 - 45, na p. 33.

¹⁴⁶ Pearson, "Mathematical Contributions to the Theory of Evolution. IX", na p. 357.

¹⁴⁷ Em suas tabulações sobre correlação fraternal, Pearson chega a um valor médio de $r = 0.4479$ e com cálculo de $\sigma = \sqrt{1 - r^2}$ encontrou um resultado igual a $r = 0.4302$.

procuraram identificar tal relação.¹⁴⁸ Consideraram o conceito da correlação, ou seja, se havia alguma relação estatisticamente significativa entre duas variáveis, representando-a através de gráficos. Calcularam o valor do coeficiente de correlação (a medida da intensidade da associação linear entre duas variáveis).

Nesta época, o conceito da correlação já era de conhecimento dos estudiosos da área, mas foi Pearson, que propôs, em seu artigo “Regression, Heredity and Panmixia” (1896), uma formulação para o mesmo. Pearson também se referiu ao “coeficiente de correlação momento-produto” que está expresso na fórmula abaixo:

$$r = \frac{n\sum xy - (\sum x)(\sum y)}{\sqrt{[n(\sum x^2) - (\sum x)^2][n\sum y^2 - (\sum y)^2]}}$$

O valor¹⁴⁹ de r é calculado com base em dados amostrais. É uma estatística amostral usada para medir o grau de correlação entre x e y , onde n representa o número de pares de dados presentes.

Tais cálculos e conceitos sobre correlação foram adotados por Pearson em seu trabalho sobre o princípio da homotipose como veremos a seguir.

3.2 ASPECTOS ESTATÍSTICOS DO PRINCÍPIO DA HOMOTIPOSE

Em seu artigo “Mathematical Contributions to the Theory of Evolution. IX” (1901), Pearson tratou de assuntos relacionados aos homótipos. Ele desejava obter esclarecimentos sobre alguns pontos:

¹⁴⁸ Para mais detalhes sobre os trabalhos de correlação, ver por exemplo, Karl Pearson, “Notes on the History of Correlation”, *Biometrika* 13 (1920): 25-45.

¹⁴⁹ O coeficiente de correlação independe das unidades de medida das variáveis x e y , pois se trata de um número adimensional e varia de -1 a 1, e não havendo relação entre x e y , temos $r = 0$.

- Se havia um maior grau de semelhança entre os homotipos de um indivíduo ou entre os homotipos de indivíduos diferentes;
- Se os homotipos de um indivíduo representam uma amostra dos homotipos da raça;
- Se entre irmãos há certo grau de semelhança e diferença (variação) entre homotipos de irmãos;
- Se a hereditariedade é um caso especial de homotipose;
- Se a correlação fraternal entre os descendentes de um mesmo progenitor é a mesma encontrada nos órgãos semelhantes indiferenciados (correlação homotípica).¹⁵⁰

Procurando encontrar respostas para essas questões, Pearson coletou grande quantidade de material do reino vegetal, elaborando no total 22 séries, como mencionamos no capítulo anterior. Dividiu o artigo sobre a homotipose em seis tópicos:

- I - Introdução;
- II – Teoria – Discussão teórica da relação da correlação fraternal e a correlação dos órgãos similares indiferenciados;
- III – Observação – Variabilidade e correlação dos órgãos similares indiferenciados do indivíduo, no qual discutiu os métodos de seleção e recolhimento do material, os processos estatísticos empregados na redução do material e a terminologia adotada: homotipos e homotipose;
- IV – Dados atuais – apresentou uma coleção de dados das folhas de várias árvores (tabelas de correlação, variabilidade e características);
- V – Cruzamento homotípico e índice de correlação;
- VI – Resumos dos resultados.

Na introdução deste trabalho, Pearson levantou questões sobre hereditariedade e comentou acerca de suas observações sobre um indivíduo que “gera um número de órgãos similares, células no sangue, pétalas de flores, folhas de árvores, tamanhos de asas”.¹⁵¹ Complementou que eles podem se diferenciar por influência de pequenos fatores ambientais e hereditários. Por exemplo, os brotos podem se diferenciar de acordo com a sua posição na flor. Inicialmente utilizou a expressão “órgãos similares indiferenciados” e depois

¹⁵⁰ Martins, “Karl Pearson, William Bateson e a Controvérsia da Homotipose”, p. 3.

¹⁵¹ Pearson, “Mathematical Contributions to the Theory of Evolution. IX”, na p. 286.

passou a chamá-los de homotipos. Pearson propôs, o “quanto a semelhança dos descendentes de um mesmo pai está relacionada à semelhança entre indivíduos de órgãos similares indiferenciados”.¹⁵² Ele comentou que para verificar uma relação teórica seria preciso dispor de uma ampla variedade de dados.¹⁵³

Ao discutir a relação entre a correlação fraternal e a correlação dos órgãos similares indiferenciados, Pearson fez cálculos que indicavam que as “características [do óvulo e do espermatozóide] certamente não poderiam ser determinadas, muito menos medidas, mas nós não temos o direito de duvidar de sua existência.”¹⁵⁴

Para demonstrar essa ideia, Pearson supôs a variável z que depende de algumas características do espermatozóide e do óvulo a partir do qual o indivíduo se desenvolveu. Determinou como variantes $x_1, x_2, x_3, \dots$ de um espermatozóide específico do qual o indivíduo se desenvolveu e $y_1, y_2, y_3, \dots$ correspondentes às variantes das características do óvulo.¹⁵⁵ Sendo:

$$z = f(x_1, x_2, x_3, \dots, y_1, y_2, y_3, \dots)$$

Segundo Pearson, f é uma função razoavelmente desconhecida. Ele assumiu as variações comparadas aos valores principais das características do espermatozóide e do óvulo, como segue:

$$z = \alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 x_3 + \dots + \beta_1 y_1 + \beta_2 y_2 + \beta_3 y_3 + \dots$$

onde α e β são independentes de x e y e definem a hereditariedade masculina e feminina.

¹⁵² Pearson, “Mathematical Contributions to the Theory of Evolution. IX”, na p. 288.

¹⁵³ Pearson acreditava que os óvulos e espermatozoides eram “órgãos similares indiferenciados”. E tais órgãos limitavam um grau de semelhança em seus descendentes, assim disse que se sua visão estivesse correta a variabilidade não é uma peculiaridade sexual, é algo peculiar à produção de “órgãos similares indiferenciados” no indivíduo, e os problemas de hereditariedade têm mostrado como a semelhança de tais órgãos é modificada pelas condições de nutrição, crescimento e geralmente meio ambiente.

¹⁵⁴ Pearson, “Mathematical Contributions to the Theory of Evolution. IX”, na p. 289.

¹⁵⁵ *Ibid.*

Pearson utilizou-se da variância populacional (σ) da característica z , atribuindo " σ_q o desvio padrão de x_q , σ'_q de y_q e deixou as variáveis r_{pq} como a correlação de x_p e x_q , r'_{pq} de y_p e y_q ."¹⁵⁶

O matemático inglês supôs que x_p e y_q , não estariam correlacionados, já que não havia seleção de um óvulo particular por um espermatozóide. Então considerou que n representaria o número de indivíduos na população e utilizou a seguinte fórmula para variância:

$$\sigma = \frac{S(z^2)}{n}$$

Onde S seria a somatória de todos os indivíduos de qualquer x e y . Pearson considerou a correlação de dois indivíduos (espermatozóide e óvulo) sendo entre eles mesmos, z_1 e z_2 seriam os valores de suas características e x', x'', y', y'' representavam as características fundamentais nos dois espermatozóides e nos dois óvulos estudados. Então, Pearson igualou:

$$z_1 = \Sigma (\alpha_p, x'_p) + \Sigma (\beta_p, y'_p)$$

$$z_2 = \Sigma (\alpha_p, x''_p) + \Sigma (\beta_p, y''_p)$$

Pearson multiplicou z_1 por z_2 , para obter o conjunto de irmãos que terão as mesmas variações (isto se eles forem do mesmo sexo). Considerou que R fosse a correlação fraternal, e tinha como:

$$S(z_1 \times z_2) = n \times \sigma \times \sigma \times R$$

No qual n é o número total dos pares de irmãos.

Apresentamos aqui, uma pequena demonstração do raciocínio de Pearson e ao final de suas formulações, concluiu que: "as características no espermatozóide e óvulo são desconhecidos e imensuráveis e não podíamos ir além."¹⁵⁷ Todo esse desenvolvimento matemático de Pearson foi para levantar

¹⁵⁶ Pearson, "Mathematical Contributions to the Theory of Evolution. IX", na p. 289.

¹⁵⁷ *Ibid*, na p. 290.

a hipótese de que a correlação entre irmãos seria igual à principal correlação entre os órgãos similares indiferenciados gerados por um indivíduo.

Neste artigo sobre a homotipose, Pearson apresentou uma coleção de dados e escolheu as folhas das árvores como amostra, pois constituíam um material de facilmente encontrado e, portanto, adequado. Apesar disso, considerou que se tratava de um trabalho complexo que envolvia tanto a coleta como a tabulação de dados relacionados às características das mesmas, que excluía as dimensões. Em suas palavras:

...poderemos contar as veias nas folhas ou folhinhas, e certificar o grau de semelhança entre as folhas da mesma árvore. A variação é em alguns casos, muito considerável, e o trabalho em formar tabelas envolvendo milhares de entradas, é muito grande.¹⁵⁸

Mesmo, Pearson vendo a complexidade de seu trabalho, na coleta e tabulação de dados, considerou que a folha seria o material mais satisfatório em sua pesquisa. Pearson descreveu detalhadamente a forma de recolhimento e verificação das características das folhas.

Em sua primeira série, Pearson e sua equipe verificaram o número de “pinnae” na folha da Ash (*Fraxinus excelsior*). Tal série consistia de 26 folhas tiradas de 109 árvores, chegou a um número de 70.850 pares. Neste caso, quase todas as árvores eram grandes e velhas e cresceram em Great e Little Hampden, Buckinghamshire. Em sua segunda série de dados, recolheram 25 folhas de cada uma das 120 árvores. Estas árvores eram jovens e estavam nos arredores de Lyme Regis em Dorsetshire. O número de pares era igual a 73.000. A terceira série consistia de 26 folhas de cada uma das 100 árvores retiradas em Monmouthshire.¹⁵⁹ Pearson apresentou uma tabela com os resultados numéricos, descrevendo o número de árvores, folhas, pares e a correlação. A tabela com os valores obtidos está reproduzida logo abaixo:

¹⁵⁸ Karl Pearson, “Mathematical Contributions to the Theory of Evolution. IX”, na p. 294.

¹⁵⁹ *Ibid*, pp.294 - 295.

Series.	Number.			Mean No. of leaflets.	S. D. of leaflets.	Correlation.
	Trees.	Leaves.	Pairs.			
Buckinghamshire	109	2834	70850	$10.1295 \pm .0214$	$1.6891 \pm .0151$	$.3743 [\pm .0109]^\dagger$
Dorsetshire . .	120	3120	78000	$9.7260 \pm .0239$	$1.9759 \pm .0169$	$.3964 [\pm .0102]$
Monmouthshire .	100	2600	65000	$9.8766 \pm .0265$	$2.0058 \pm .0188$	$.4047 [\pm .0111]$
Mean of series .	—	—	—	9.9107	1.8903	.3918

Tabela 3.2 : Semelhança das folhas Ash

(Fonte: Karl Pearson, “Mathematical Contributions to the Theory of Evolution. IX. (1901), p. 295)

Nessa tabulação, Pearson apresentou 8.554 folhas compostas. Contou-as e distribuiu os 213.850 pares nas tabelas simétricas¹⁶⁰. Para o matemático inglês, esses dados forneceram uma correlação de 0.3918, o que era razoavelmente similar ao valor teórico de 0.4000 dado pela “Lei da herança ancestral” para o grau de semelhança entre os pares de irmãos.¹⁶¹

Pearson também tabulou o número de veias nas folhas da faia e do castanheiro espanhol. Detalhou a coleta (ao acaso, ao redor da árvore; no outono, etc.) e o modo pelo qual contou o número de veias nestas folhas (veias perto da ponta da folha, uso de lentes, observação da folha em vários ângulos, etc.).¹⁶² Calculou a média das veias do castanheiro espanhol, chegou a um número modal das folhas, e comparou a variabilidade das raças (*Castanea vulgaris* e *Fraxinus excelsior*).¹⁶³ Contou o número de espinhos do azevinho, as listras estigmáticas nas cápsulas de sementes (flor e broto) das papoulas, obtendo assim a maior tabela de correlação em seu trabalho, com a papoula *Chelsea* tendo uma entrada de 197.478 pares.

Ao tratar das características das folhas coletadas, Pearson comentou que havia uma diferença entre os critérios adotados por um botânico e por um estatístico na escolha das características a serem estudadas. Ele comentou:

¹⁶⁰ As tabelas com as distribuições estão em anexo.

¹⁶¹ Pearson, “Mathematical Contributions to the Theory of Evolution. IX”, na p. 295.

¹⁶² *Ibid*, na p. 300.

¹⁶³ *Ibid*, na p. 302.

Devo confessar que há divergência entre os pontos de vistas do botânico e do estatístico. Para o último uma característica é boa ou ruim se ela propicia facilidades para uma medição ou enumeração razoavelmente fácil. Primeiro, ele tem que procurar essas características como condição indispensável, e depois questionar o quanto elas ocorrem nos órgãos similares indiferenciados. Por outro lado, o biólogo proverá uma lista de tais características na vida animal ou vegetal e não naquelas que estão disponíveis para medição e a contagem.¹⁶⁴

Além de explicar os critérios que utilizou na seleção das características estudadas, Pearson comentou sobre as dificuldades que encontrou em suas investigações. Precisou utilizar muitas vezes instrumentos como lupa, microscópio, por exemplo. Mesmo sendo auxiliado por uma equipe empenhada neste estudo, em alguns momentos trabalhou sozinho na contagem de certas séries.

3.3 RESULTADOS DA CORRELAÇÃO

Ao concluir seus estudos sobre a correlação Pearson comentou: “Encontrei certa dificuldade em resumir e comparar os resultados que obtive sobre correlação fraternal com aqueles obtidos pelos meus colegas”¹⁶⁵ Mesmo com tais dificuldades em relação à análise dos resultados, Pearson incluiu todos os resultados homotípicos. Assim, elaborou a tabela 3.3 contendo o resumo de todos os resultados que havia obtido no estudo sobre correlação homotípica e a tabela 3.4 sobre correlação fraternal. Nessas análises, Pearson preocupou-se em mostrar a veracidade de seu trabalho e que não descartou as influências que poderiam fazer com que valores incorretos fossem incluídos nas tabelas. Assim, esclareceu que havia lidado com uma grande variedade de

¹⁶⁴ Pearson, “Mathematical Contributions to the Theory of Evolution. IX.”, pp. 324 - 325.

¹⁶⁵ *Ibid*, na p. 355.

raças e características, e que qualquer influência externa iria equilibrar a média do conjunto. Ele assim se expressou:

O montante das causas atuantes e perturbadoras não pode ser medido (...). Só espero ter tomado uma gama bastante ampla de raças e características e que as influências que tendem a obscurecer a correlação homotípica por um lado, aumentando-a e por outro, reduzindo-a sejam equilibradas, e que a média do conjunto esteja próxima do verdadeiro valor da correlação homotípica.¹⁶⁶

Ao analisar os valores encontrados na tabela 3.4, Pearson se deu conta que o valor da correlação para *Daphnia* era muito alto (0.6934) e o valor da longevidade do homem (0.2602) estava muito abaixo da verdadeira correlação fraternal.¹⁶⁷ Verificou que a média da correlação fraternal (Tabela 3.4) teve valor igual a 0.4479 e para a correlação homotípica (Tabela 3.3) a média foi de 0.4570. Concluiu que os valores médios das duas correlações estavam dentro dos limites dos erros prováveis, sendo considerados praticamente iguais.¹⁶⁸ Nesse sentido, ele assim se expressou:

... eu gostaria de pedir ao leitor, uma análise das duas tabelas lado a lado, (...) verificando que há provas suficientes a partir de uma gama bastante ampla de caracteres (...) e de espécies para mostrar que tanto a correlação homotípica ou fraternal flutuam sobre um valor médio entre 0.4 e 0.5.¹⁶⁹

Pearson comentou sobre sua dificuldade em resumir os vários fatores que influenciaram as formas de vida e por isso se preocupou somente em

¹⁶⁶ Pearson, "Mathematical Contributions to the Theory of Evolution. IX", na p. 355.

¹⁶⁷ A explicação de Pearson para tais resultados foi que no primeiro caso, as mães eram em número reduzido, e no outro caso, a redução dos valores se deu pela taxa não seletiva de morte.

¹⁶⁸ Pearson, "Mathematical Contributions to the Theory of Evolution. IX", na p. 358.

¹⁶⁹ *Ibid.*

medir o fator que se estava investigando. Em relação aos valores obtidos, Pearson explicou que se os resultados da homotipose fossem exatamente 0.45, ele estaria inclinado a desconfiar de qualquer uma das séries de resultados, em cada um que apresentasse uma resposta exatamente igual a 0.45.¹⁷⁰

Abaixo reproduzimos as tabelas com os resultados da correlação homotípica e correlação fraternal:

¹⁷⁰ Pearson, "Mathematical Contributions to the Theory of Evolution. IX", na p. 358.

Race.	Character.	Per-centage variation.	Corre-lation.	Remarks.
Mushroom, Hampden . . .	Lengths of gills .	50.92	.8607	All these results introduce a correlation due to stages of growth and accordingly are not included in the determination of means.
" " . . .	Breadths of gills .	67.67	.7363	
" " . . .	Lengths and breadths of gills	—	.6275	
Wild Ivy, mixed localities .	Lengths of leaves .	82.73	.5618	
" " . . .	Breadths of leaves	84.56	.5332	
" " . . .	Lengths and breadths of leaves	—	.5157	
(i.) Ceterach, Somersetshire	Lobes on fronds .	77.57	.6311	Said to be largely affected by growth and environment.
(ii.) Hartstongue, Somersetshire	Sori on fronds .	77.64	.6303	
(iii.) Shirley Poppy, Chelsea .	Stigmatic bands .	78.86	.6149	
(iv.) English Onion, Hampden	Veins in tunics .	79.18	.6108	Much selected in transit. Possibly slightly hetero- geneous.
(v.) Holly, Dorsetshire . . .	Prickles on leaves	80.11	.5985	Heterogeneous.
(vi.) Spanish Chestnut, mixed	Veins in leaves .	80.65	.5913	
(vii.) Beech, Buckinghamshire	Veins in leaves .	82.17	.5699	
(viii.) <i>Papaver Rhoeas</i> , Hampden	Stigmatic bands .	82.71	.5620	Possibly influenced by indi- vidual growth.
(ix.) Mushroom, Hampden .	Gill indices . . .	83.58	.5490	
(x.) <i>Papaver Rhoeas</i> , Quantocks	Stigmatic bands .	84.59	.5333	
(xi.) Shirley Poppy, Hampden	Stigmatic bands .	85.18	.5238	All from one field.
(xii.) Spanish Chestnut, Buck- inghamshire	Veins in leaves .	88.51	.4655	
(xiii.) Broom, Yorkshire . . .	Seeds in pods . .	90.96	.4155	
(xiv.) Ash, Monmouthshire .	Leaflets on leaves .	91.44	.4047	From two localities and pos- sibly slightly influenced by differentiation.
(xv.) <i>Papaver Rhoeas</i> , Lower Chilterns	Stigmatic bands .	91.66	.3997	
(xvi.) Ash, Dorsetshire . . .	Leaflets on leaves .	91.81	.3964	
(xvii.) Ash, Buckinghamshire .	Leaflets on leaves .	92.73	.3743	Differentiation of organs due to position on stem. Principally spread from one clump by stolons.
(xviii.) Holly, Somersetshire .	Prickles on leaves	93.12	.3548	
(xix.) Wild Ivy, mixed localities	Leaf indices . . .	96.21	.2726	
(xx.) <i>Nigella Hispanica</i> , Slough	Segments of seed- capsules	98.18	.1899	Members really different in morphological origin.
(xxi.) <i>Malva Rotundifolia</i> , Hampden	Segments of seed- vessels	98.32	.1827	
(xxii.) Woodruff, Buckingham- shire	Members of whorls	98.49	.1733	
Mean of 22 cases . . .	—	87.44	.4570	—

Tabela 3.3 : Resultados Gerais para Correlação Homotípica.
(Fonte: Karl Pearson, Mathematical Contributions to the Theory of Evolution,
1901, p.356)

Race.	Sex.	Character.	Source of material.	No. of cases.	Reduced by	Correlation.	Remarks.
(i.) Daphnia	♀ & ♀	Length of spine	ERNEST WARREN . . .	330	K. PEARSON . . .	.6934	Probably much too high, owing to heterogeneity introduced by the selection of a few mothers only.
(ii.) Horse	♀ & ♀	Coat-colour.	WEATHERBY'S Studbooks . .	1000	K. PEARSON, L. BRAMLEY-MOORE, and A. LEE	.6928	Probably much too high, owing to heterogeneity introduced by the use of comparatively few sires.
(iii.) "	♂ & ♂	"	"	1000		.6232	
(iv.) "	♂ & ♀	"	"	1000		.5827	
(v.) Man.	♀ & ♀	Forearm.	PEARSON, family data . .	441	A. LEE . . .	.5424	One pair only from each family.
(vi.) Hound (Basset)	mixed	Coat-colour.	GALTON, from studbook . .	—	K. PEARSON and A. LEE	.5257	All members of litter without regard to sex.
(vii.) Man.	♂ & ♂	Eye-colour . .	GALTON, family data . .	1500	K. PEARSON . . .	.5169	All possibly pairs in family taken.
(viii.) "	♀ & ♀	Cephalic index	FRANZ BOAS, N. A. Indians	—	C. FAWCETT . . .	.4890	Paternity doubtful.
(ix.) "	♂ & ♀	Eye-colour . .	GALTON, family data . .	1500	K. PEARSON . . .	.4615	See remark to (vii.).
(x.) "	♀ & ♀	"	"	1500	"	.4463	"
(xi.) "	♀ & ♀	Stature . . .	"	595	"	.4436	"
(xii.) "	♂ & ♂	"	"	605	"	.3913	"
(xiii.) "	♂ & ♂	Cephalic index	FRANZ BOAS, N. A. Indians	—	C. FAWCETT . . .	.3790	See remark to (viii.).
(xiv.) "	♂ & ♀	Stature . . .	GALTON, family data . .	1181	K. PEARSON . . .	.3754	See remark to (vii.).
(xv.) "	♂ & ♀	Cephalic index	FRANZ BOAS, N. A. Indians	—	C. FAWCETT . . .	.3400	See remark to (viii.).
(xvi.) "	♀ & ♀	Longevity . .	Quaker records . . .	1050	M. BEETON . . .	.3323	Reduced below true value by non-selective deaths.
(xvii.) "	mixed	Temper . . .	GALTON, family data . .	1294	K. PEARSON . . .	.3167	Character very indefinite, and difficult to estimate.
(xviii.) "	♂ & ♂	Longevity . .	Peerage records . . .	1000	M. BEETON . . .	.2602	See remark to (xvi.).
(xix.) "	♂ & ♀	"	Quaker records . . .	1947	"	.1973	"
—	—	—	—	—	Mean of 19 series	.4479	—

Tabela 3.4: Resultados Gerais para Correlação Fraternal

(Fonte: Karl Pearson, "Mathematical Contributions to the Theory of Evolution", 1901, p.357)

Após analisar e comparar os resultados das correlações obtidas nas tabelas, Pearson apresentou sua visão acerca da hereditariedade. Para ele, a hereditariedade seria somente um caso especial da homotipose¹⁷¹, no qual “qualquer multiplicação das células é homotípica e denota um grau dado de variação e um grau de semelhança.”¹⁷²

Pearson deduziu a relação expressa acima a partir dos resultados das tabelas. A intensidade da homotipose em todo o reino vegetal situa-se entre 0.4 e 0.5. Este foi também o valor médio que ele encontrou para a correlação fraternal. Em suas palavras:

Podemos concluir que a hereditariedade é realmente apenas uma fase maior da homotipose. No fundo, é apenas uma parte do princípio de que quando um indivíduo produz órgãos indiferenciados, estes órgãos não são exatamente os mesmos, mas apresentam uma determinada intensidade de variação e um determinado grau de semelhança. Quando associamos hereditariedade com reprodução sexual, estamos considerando apenas o resultado da homotipose (variação e semelhança) entre espermatozóides individuais e óvulos individuais. Tal homotipose conduz a uma semelhança dos indivíduos resultantes do zigoto que, provavelmente, apresentam o mesmo valor médio de correlação homotípica. Assim, se a continuidade das células do embrião entre pais e seus descendentes ocorrer, nós não enfrentaremos mais o problema da hereditariedade, mas o da homotipose, e isto poderia ser reduzido ao simples fato da multiplicação de brotos e células.¹⁷³

¹⁷¹ Pearson discutiu que não se podia esperar, teoricamente, mais variabilidade em espécies reproduzidas sexualmente, do que em espécies reproduzidas assexuadamente. Ele aderiu a esta crença durante muitos anos, mas depois foi um dos primeiros a deduzir que em espécies reproduzidas sexualmente a combinação genética, prevista pela hereditariedade mendeliana, poderia providenciar um vasto número de variantes genéticas.

¹⁷² Pearson, “Mathematical Contributions to the Theory of Evolution. IX”, na p. 294.

¹⁷³ *Ibid*, na p. 359.

Segundo Provine, se este princípio fosse verdadeiro, a teoria da homotipose de Pearson teria sido uma surpreendente contribuição à biologia, porém seus fundamentos biológicos eram fracos.¹⁷⁴

No artigo sobre o princípio da homotipose, Pearson prosseguiu enfatizando a variabilidade dos órgãos similares indiferenciados. Relatou que a homotipose envolve não somente certo grau de semelhança no grupo de homotipos, mas também certo grau de variação. Ao calcular a variação percentual do indivíduo, Pearson utilizou o desvio padrão do conjunto, pois para ele é uma medida razoável de variabilidade. Mostrou tais valores como na tabela 3.3 na coluna intitulada variação percentual (percentage variation). Em que excluindo¹⁷⁵ os valores correspondentes ao comprimento e largura dos cogumelos em que a variação percentual está abaixo de 77, Pearson verificou que a média da variação do indivíduo está entre 87 e 88 por cento daquela raça e que os órgãos similares indiferenciados apresentados pelo indivíduo tem números exatos de 80 a 90 por cento de variação de tais órgãos no grupo. Ao apresentar estes dados, Pearson fez uma crítica implícita aos biólogos: “Agora, isto me parece muito instrutivo, quando consideramos as declarações feitas por alguns escritores que teorizam sobre variação, ao invés de realmente medi-la.”¹⁷⁶ Questionou se a variação não seria insignificante no caso da germinação.

Pearson apresentou uma tabela que descrevia os resultados da variabilidade das séries estudadas (dado pelas vinte e duas séries) e a relacionou ao posicionamento da tabela de correlação, como segue:

¹⁷⁴ William B. Provine, “The origins of theoretical population genetics”, University of Chicago Press, 2001, p. 60.

¹⁷⁵ Pearson relatou que nesses casos o estágio de crescimento é muito importante no resultado da variação percentual.

¹⁷⁶ Pearson, “Mathematical Contributions to the Theory of Evolution. IX”, na p. 359.

Race.	Character.	Coefficient of variation.	Position in correlation table.
Hartstongue, Somersetshire	Sori on fronds	41.96	(ii.)
Broom, Yorkshire	Seeds in pods	36.78	(xiii.)
Holly, Dorsetshire	Prickles on leaves	26.29	(v.)
<i>Nigella Hispanica</i> , Slough	Segments in capsules	19.39	(xx.)
Holly, Somersetshire	Prickles on leaves	18.74	(xviii.)
Ash, Dorsetshire	Leaflets on leaf	18.65	(xvi.)
Ash, Monmouthshire	" "	18.57	(xiv.)
Mushroom, Buckinghamshire	Gill-indices	18.28	(ix.)
Ceterach, Somersetshire	Lobes on fronds	18.25	(i.)
<i>Papaver Rhæas</i> , top of Chilterns	Stigmatic bands	17.81	(viii.)
Wild Ivy, mixed	Leaf-indices	17.77	(xix.)
<i>Papaver Rhæas</i> , Quantocks	Stigmatic bands	17.66	(x.)
English Onion, Hampden	Veins in tunics	17.43	(iv.)
Spanish Chestnut, mixed	Veins in leaves	15.72	(vi.)
Ash, Buckinghamshire	" "	15.46	(xvii.)
<i>Papaver Rhæas</i> , Lower Chilterns	Stigmatic bands	15.27	(xv.)
Shirley Poppy, Hampden	" "	15.17	(xi.)
Spanish Chestnut, Buckinghamshire	Veins in leaves	14.31	(xii.)
Shirley Poppy, Chelsea	Stigmatic bands	13.99	(iii.)
Woodruff, Buckinghamshire	Members in whorls	13.46	(xxii.)
Beech, Buckinghamshire	Veins in leaves	10.77	(vii.)
<i>Malva Rotundifolia</i> , Hampden	Segments in seed-vessels	7.80	• (xxi.)
Mushroom, Buckinghamshire	Lengths of gills	21.00	—
" " " "	Breadth of gills	23.42	—
Wild Ivy, mixed " "	Lengths of leaves	30.94	—
" " " "	Breadths of leaves	34.34	—
Mean of first 22 cases	—	18.62	—
Mean of all 26 cases	—	19.97	—

Tabela 3.5 : Resultados para variabilidade

(Fonte: Karl Pearson, "Mathematical Contributions to the Theory of Evolution", 1901, p.361)

Esta tabela 3.5 mostra a variação percentual de uma característica específica, Pearson não comparou características diferentes em espécies diferentes. Isso pode ser observado a partir dos resultados na coluna que apresenta os coeficientes de variação. Pearson verificou uma série razoavelmente contínua de valores entre 7.80 a 41.96, tendo valor médio de 19.¹⁷⁷

¹⁷⁷ Para Pearson este resultado médio da variação está de acordo com os resultados obtidos em séries mais amplas no reino vegetal.

Ao examinar a quarta coluna da tabela 3.5, que relaciona a correlação homotípica correspondente, Pearson concluiu que não havia relação entre a intensidade da variabilidade e da homotipose. Em suas palavras:

Não foi encontrada nenhuma relação entre a intensidade da variabilidade e a homotipose nas vinte e duas séries estudadas. E nenhuma relação observada entre a simplicidade relativa do organismo e a intensidade, quer em sua variabilidade ou em sua correlação homotípica. Considerando hereditariedade como um caso da homotipose, ela é um fator primário das formas vivas, uma condição para a evolução da vida pela seleção natural, e não produto de tal seleção. Se o cogumelo, a papoula e a faia mostram aproximadamente uma homotipose semelhante, parece-me impossível considerá-la um fator de vida que aumenta com o avançar da evolução.¹⁷⁸

Para Pearson, a ordem homotípica não pode ser comparada de nenhum modo à ordem de variação. Assim, ele analisa nesta tabela 3.5, que nas onze primeiras séries a variabilidade média é de 22.95 e nas últimas onze séries é de 14.28, sendo um desvio da média de 18.62 de todas as vinte e duas séries. Mas a correlação homotípica média das primeiras onze séries é 0.4559, e das últimas onze séries é 0.4581, e nenhuma delas exibiu uma maior diferença em relação à média de 0.4570 (obtida na tabela). A partir desses resultados, Pearson afirmou que “seria impossível afirmar que as variabilidades das espécies decresceram, e suas homotiposes cresceram.”¹⁷⁹ E acrescentou:

Isso não me parece que as espécies em relação ao topo sejam mais primitivas e aquelas da parte inferior (da tabela) sejam mais complexas. (...) Sou forçado a concluir que minha tabela não dá nenhum suporte, a qualquer ponto de vista que a variabilidade no reino vegetal é maior nos organismos mais simples e primitivos. (...)

¹⁷⁸ Pearson, “Mathematical Contributions to the Theory of Evolution. IX”, na p. 363.

¹⁷⁹ *Ibid*, na p. 360.

Não vejo relação entre a posição de um organismo na descendência evolucionária e a intensidade da variabilidade ou sua homotipose.¹⁸⁰

Pearson justificou tais resultados relatando que suas estatísticas cobriram somente uma pequena parte do reino vegetal, porém segundo ele, elas representam o “único” material até o presente, (...) obtidos a partir de fatos quantitativos e não adquiridos de uma estimativa geral.¹⁸¹

Neste estudo Pearson chegou às seguintes conclusões:

- i) Os homotipos apresentam certo grau de variação combinada com certo grau de semelhança (correlação homotípica);
- ii) A correlação homotípica (testada nas 22 séries do reino vegetal) tem valor médio entre 0.4 e 0.5;
- iii) A variabilidade individual (característica numa série de homotipos) é de 80 a 90 por cento da variabilidade desta característica na espécie como um todo;¹⁸²
- iv) A correlação fraternal se igualaria à correlação homotípica;
- v) Se (iii) for verdadeira, a hereditariedade é uma fase única do princípio da homotipose e que (ii) o valor numérico de sua constante pode ser encontrado nesse princípio;
- vi) Nenhuma relação pode ser encontrada entre a intensidade da variabilidade e a homotipose das vinte e duas séries utilizadas.

Percebemos que Pearson baseou-se nos resultados estatísticos (valores numéricos) para justificar suas concepções acerca da hereditariedade e evolução, e que tais resultados pouco mudaram suas hipóteses iniciais. Seus dados, cálculos e tabelas descreveram corretamente as características verificadas nos homotipos estudados, porém sua preocupação foi relacionar hereditariedade e homotipose utilizando a biometria para justificar tais conceitos biológicos e hipóteses.

¹⁸⁰ Pearson, “Mathematical Contributions to the Theory of Evolution. IX”, na p.362.

¹⁸¹ *Ibid.*

¹⁸² Assim, para Pearson, a variabilidade é um fator primário de formas vivas, e não depende de, em nenhuma hipótese, de reprodução sexual.

CAPÍTULO 4

CONSIDERAÇÕES FINAIS

O objetivo inicial desta dissertação, conforme aparece na introdução consistia em descrever algumas concepções estatísticas de Karl Pearson, no período considerado (1899-1906), relacionadas ao princípio da homotipose, além da sua coleta de materiais e tabulações presentes em seu artigo. Além disso, elucidar quais as concepções biológicas em que ele se baseou em seus estudos, que abrangiam alguns aspectos da seleção natural de Darwin, através de um estudo estatístico de populações de plantas e animais.

Ao que tudo indica as primeiras anotações de Karl Pearson dedicadas à matemática estatística aparecem por volta de 1893 (com as publicações de “The Geometry of Chance” e “Asymmetrical frequency curve”) e, embora já houvesse publicado vários trabalhos (mais de 20 trabalhos) antes de 1901, seus principais resultados pertenceram ao século XX. Podemos dizer que, Pearson difundiu a correlação através de um grande número de tabelas estatísticas, trabalhando a estimativa de seus parâmetros pela observação. Desta forma, contribuiu para uma tendência da matemática estatística, (na virada do século XIX e início do século XX) em deixar o método de determinar “valores verdadeiros” de quantidades observadas (na clássica teoria dos erros) e estimar os parâmetros de distribuição. Isso possibilitou aos estatísticos da época levantarem problemas concretos de forma mais clara e criar o relevante ramo da matemática estatística.

Como podemos perceber nas décadas de 1890 e 1900, Pearson e sua equipe investigaram uma grande distribuição amostral de uma população, e tomaram como objetivo a compreensão da variação, hereditariedade e seleção no reino animal e vegetal. Ao que tudo indica, tais estudos foram baseados em números (estatisticamente) grandes de espécimes, pois eles não estavam preocupados com a análise de dados de experimentos em pequena escala. Eles desenvolviam uma teoria estatística com tabelas de funções matemáticas e métodos gráficos.

No trabalho sobre homotipose (1901), Pearson observou a existência de órgãos similares indiferenciados, como as folhas de uma mesma árvore. Ele sentiu necessidade em introduzir novas terminologias (homotipose) para descrever suas ideias. Coletou uma grande variedade de espécies do reino vegetal, verificando as características que fossem fáceis de medir, tabular e correlacionar. Segundo Porter, os estudos de Pearson sobre correlação foram a chave para a biometria. Esta, apoiada nas concepções da seleção natural, requeria uma medida da interação entre órgãos similares ou características que tendiam a variar.¹⁸³

Nestes estudos, Pearson assumiu que os espermatozóides e óvulos eram órgãos similares indiferenciados. Porém, em contrapartida, as concepções de Mendel, indicavam que os gametas eram diferenciados.¹⁸⁴

Para Pearson, o princípio da homotipose era fundamental na natureza e que de algum modo poderia ser “a fonte da hereditariedade”.¹⁸⁵ Segundo Provine, a teoria de Mendel foi claramente contraditória à teoria da homotipose de Pearson, e na elaboração de seu argumento teórico, Pearson baseou-se em uma suposição duvidosa, ao igualar correlação fraternal à correlação homotípica. Nessa suposição e nos valores obtidos nas tabelas, Pearson acreditou ter provado que a variação de órgãos semelhantes indiferenciados em um indivíduo era o mesmo fenômeno que a variação entre irmãos.¹⁸⁶ Para Kolmogorov, os trabalhos de Pearson foram caracterizados por um nível teórico razoavelmente baixo, uma circunstância que impediu o seu reconhecimento e maior desenvolvimento fora da Inglaterra.¹⁸⁷

Verificamos que em seu artigo sobre o princípio da homotipose, Pearson não se preocupou em desenvolver uma teoria ou técnicas estatísticas, ele tinha em mente as aplicações dos métodos estatísticos para a resolução dos problemas que envolviam hereditariedade e evolução. Pearson apresentou além dos métodos estatísticos algumas hipóteses aos princípios biológicos

¹⁸³ Theodore M. Porter. *Karl Pearson: the scientific life in a statistical age*. Princeton University Press, New Jersey: 1953, p. 257.

¹⁸⁴ Gregor Mendel, “Experiments in plant hybridization”, *Natural History Society* (1865): 1-39, na p.35.

¹⁸⁵ Lilian Al-Chueyr Pereira Martins, “Karl Pearson, William Bateson e a controvérsia da homotipose”, *Episteme* (2006):1-9, na p. 3.

¹⁸⁶ William Provine, *The origins of theoretical population genetics*. Chicago: The University of Chicago Press, 2001, pp.60 - 61.

¹⁸⁷ Kolmogorov, *Mathematics of the 19th century*, p. 275.

referentes aos órgãos similares indiferenciados. Podemos concluir que, mesmo Pearson adotando conceitos estatísticos coerentes para a explicação de seu princípio, baseou-se em conceitos biológicos equivocados.

ANEXO

TABLE I.—(i.) Buckinghamshire Ashes.

Number of Pinnae on First Leaf.

	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	15.	16.	Totals.
3	6	0	6	6	15	0	6	0	33	3	—	—	—	—	75
4	0	0	0	0	0	0	0	0	0	0	—	—	—	—	0
5	6	0	18	16	74	14	155	18	94	3	2	—	—	—	400
6	6	0	16	24	96	25	213	29	113	3	0	—	—	—	525
7	15	0	74	96	716	211	2168	235	1385	34	87	0	4	—	5025
8	0	0	14	25	211	50	732	82	485	17	56	0	3	—	1675
9	6	0	155	213	2168	732	8970	1305	7361	278	756	11	20	—	21975
10	0	0	18	29	235	82	1305	278	1648	69	213	6	17	—	3900
11	33	0	94	113	1385	485	7361	1648	13558	838	2825	57	103	—	28500
12	3	0	3	3	34	17	278	69	838	90	326	19	20	—	1700
13	—	—	2	0	87	56	756	213	2825	326	1896	88	176	—	6425
14	—	—	—	—	0	0	11	6	57	19	88	16	28	—	225
15	—	—	—	—	4	3	20	17	103	20	176	28	54	—	425
16	—	—	—	—	—	—	—	—	—	—	—	—	—	—	0
Totals	75	0	400	525	5025	1675	21975	3900	28500	1700	6425	225	425	0	70850

Number of Pinnae on Second Leaf.

Fonte: Karl Pearson, "Mathematical Contributions to the Theory of Evolution.

IX. On the Principle of Homotyposis and Its Relation to Heredity, to the Variability of the Individual, and to the of that Race. Part I. Homotypos in the Vegetable Kingdom", p.297.

TABLE II.--(ii.) Dorsetshire Ashes.
Number of Pinnae on First Leaf.

	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	15.	16.	Totals.
3	2	0	2	0	23	1	37	2	24	5	4	—	—	—	100
4	0	0	0	0	0	0	0	0	0	0	0	—	—	—	0
5	2	0	170	36	506	129	786	108	324	6	28	3	2	—	2100
6	0	0	36	22	114	39	245	62	186	11	31	0	4	—	750
7	23	0	506	114	2298	465	3726	641	1816	83	210	6	12	—	9900
8	1	0	129	39	465	454	782	338	536	31	88	2	10	—	2875
9	37	0	786	245	3726	782	9428	1662	6172	308	793	13	23	—	23975
10	2	0	108	62	641	338	1662	820	1554	103	360	9	40	1	5700
11	24	0	324	186	1816	536	6172	1554	8806	676	2421	62	192	6	22775
12	5	0	6	11	83	31	308	103	676	114	417	16	28	2	1800
13	4	0	28	31	210	88	793	360	2421	417	2276	120	238	14	7000
14	—	—	3	0	6	2	13	9	62	16	120	10	34	0	275
15	—	—	2	4	12	10	23	40	192	28	238	34	140	2	725
16	—	—	—	—	—	—	—	1	6	2	14	0	2	0	25
Totals...	100	0	2100	750	9900	2875	23975	5700	22775	1800	7000	275	725	25	78000

Number of Pinnae on Second Leaf.

Fonte: Karl Pearson, "Mathematical Contributions to the Theory of Evolution.

IX. On the Principle of Homotyposis and Its Relation to Heredity, to the Variability of the Individual, and to the of that Race. Part I. Homotypos in the Vegetable Kingdom", p.298.

TABLE III.—(iii.) Monmouthshire Ashes.

Number of Pinnae on First Leaf.

	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	15.	16.	Totals.
3	—	—	4	2	9	1	8	0	1	—	—	—	—	—	25
4	—	2	0	5	19	3	66	8	21	1	—	—	—	—	125
5	4	0	34	18	208	36	366	47	269	24	38	6	—	—	1050
6	2	5	18	10	106	21	251	32	138	2	15	0	—	—	600
7	9	19	208	106	1450	280	2594	283	1593	132	277	9	14	1	6975
8	1	3	36	21	280	60	464	86	365	19	38	1	1	0	1375
9	8	66	366	251	2594	464	8022	1227	6468	430	950	24	28	2	20900
10	0	8	47	32	283	86	1227	222	1344	94	216	2	13	1	3575
11	1	21	269	138	1593	365	6468	1344	9292	841	1924	46	93	5	22400
12	—	1	24	2	132	19	430	94	841	138	366	12	16	0	2075
13	—	—	38	15	277	38	950	216	1924	366	1388	45	131	12	5400
14	—	—	6	0	9	1	24	2	45	12	45	2	3	0	150
15	—	—	—	—	14	1	28	13	93	16	131	3	22	4	325
16	—	—	—	—	1	0	2	1	5	0	12	0	4	0	25
Totals...	25	125	1050	600	6975	1375	20900	3575	22400	2075	5400	150	325	25	65000

Number of Pinnae on Second Leaf.

2 2 2

Fonte: Karl Pearson, "Mathematical Contributions to the Theory of Evolution.

IX. On the Principle of Homotyposis and Its Relation to Heredity, to the Variability of the Individual, and to the of that Race. Part I. Homotypos in the Vegetable Kingdom", p.299.

BIBLIOGRAFIA

- BATESON, William. "Heredity, Differentiation, and Other Conceptions of Biology: A consideration of Professor Karl Pearson's Paper 'On the Principle of Homotyposis', *Royal Society* 69 (1901): 1- 404.
- CAMP, Burton H. "Karl Pearson and Mathematical Statistics", *Journal of The American Statistical Association* (184, 1933): p.395-401.
- DE MORGAN, Augustus. *An Essay on Probabilities, and on Their Application to Life Contingencies and Insurance Offices*. London: Longman, Orme, Brown, Green & Longmans, and John Taylor, 1838.
- EDGEWORTH, Francis Ysidro. "Correlated averages", *Philosophical Magazine* 34 (1892): 190-204.
- EDWARDS, Anthony William Fairbank. "R.A. Fisher on Karl Pearson", *Notes and Records of the Royal Society of London* 48 (1, 1994): 97-106.
- EISENHART, Churchill. "Pearson, Karl". In: Charles Coulston Gillispie, ed. *Dictionary of scientific biography*. Vol. 9 (New York: Charles Scribner's sons, 1981): 447 – 473.
- GALTON, Francis. "The average contribution of each several ancestor to the total heritage of the offspring". *Proceeding of the Royal Society* 61 (1897): 401 - 413.
- _____. "Typical Laws of Heredity". *Nature* 15 (1877): 285 – 297.
- _____. *Natural inheritance*. London: Macmillan & Co, 1889.
- HILTS, Victor. "Statistics and Social Science," in R. Giere and R. Westfall (eds.). *Foundations of Method: The Nineteenth Century* Bloomington: Indiana University Press (1973), pp. 206-233.
- KATZ, Victor J. *A History of Mathematics: an introduction*. Reading, MA: Addison-Wesley, 1998.
- KOLMOGOROV, Andrei Nikolaevich & Adolf-Andrei Pavlovich Yushkevich. *Mathematics of the 19th century*. Berlin: Birkhauser, 1992.

- MAGNELLO, Eileen. "Karl Pearson and the Origins of Modern Statistics: An Elastician becomes a Statistician". *The Rutheford Journal* 2 (2006-2007): 1-12.
- MARTINS, Lilian Al-Chueyr Pereira. "Weldon, Pearson, Bateson e a controvérsia mendeliano-biometricista: uma disputa entre evolucionista". *Filosofia Unisinos* 8 (2,2007): 170-190.
- _____. "Bateson e o programa de pesquisa mendeliano". *Episteme*, 2006.
- _____. "Karl Pearson, William Bateson e a controvérsia da homotipose", *Episteme*, 2006.
- _____. "William Bateson: da evolução à genética". *Episteme* (8, 1999): 67-88.
- MENDEL, Gregor. "Experiments in plant hybridization". *Natural History Society* (1865): 1-39.
- NORDMANN, Alfred. "Darwinians at war. Bateson's place in histories of Darwinism." *Synthese* 91 (1992): 53-72.
- PEARSON, Egon S. "Some reflexions on continuity in the development of mathematical statistics, 1885-1920." *Biometrika* 54 (1967): 341-355.
- PEARSON, Karl. "Notes on the history of correlation". *Biometrika* 13 (1920): 25 – 45.
- _____. "Mathematical contributions to the theory of evolution. III. Regression, heredity and panmixia". *Philos. Trans. Royal Soc. London* 187 (1896): 253 - 318.
- _____. "Mathematical Contributions to the Theory of Evolution. IX. On the Principle of Homotyposis and Its Relation to Heredity, to the Variability of the Individual, and to that of the Race. Part I. Homotypos in the Vegetable Kingdom". *Philosophical Transactions of the Royal Society of London* 197 (1901): 285 - 379.
- _____. "On the fundamental conceptions of Biology". *Biometrika* 1 (1902): 320 – 323.
- _____. "Walter Frank Raphael Weldon. 1860–1906". *Biometrika* 5 (1, 1906): 1–52.
- PEARSON, Karl & Marion Radford. "On Differentiation and Homotyposis in the Leaves of *Fagus sylvatica*". *Biometrika* 3 (1, 1904): 104-107.
- POINCARÉ, Henri. *Calcul des probabilités*. Paris : Gauthier-Villars, 1896.

- PORTER, Theodore M. *Karl Pearson: the scientific life in a statistical age*. New Jersey: Princeton, 1953.
- PORTER, Theodore M. *The English biometric tradition*. Companion Encyclopedia of the History and Philosophy of the Mathematical Sciences, vol. 2, London: Grattan-Guinness (1994): 1335 – 1362.
- PROVINE, William. *The origins of theoretical population genetics*. Chicago: The university of Chicago Press, 2001.
- SIMON, Gary A. *Estatística aplicada*. Porto Alegre: Atlas, 2000.
- SPIEGEL, Murray R. *Probabilidade e estatística*. São Paulo: Pearson, 1978.
- STIGLER, Stephen M. *Statistics on the table: The history of statistics concepts and methods*. Cambridge, MA: Harvard University Press, 2000.
- STRUIK, Dirk J. *História concisa das Matemáticas*. 2ª Ed. Lisboa: Ed. Gradiva, 1992.
- WALPOLE, Ronald E. [ET AL.]. *Probabilidade e estatística para engenharia e ciências*. São Paulo: Pearson Prentice Hall, 2009.