



UNIVERSIDADE
ESTADUAL DE LONDRINA

UNIVERSIDADE ESTADUAL DE LONDRINA

CCE - Centro de Ciências Exatas

DSTA - Departamento de Estatística

Apostila de Estatística

Prof. M.e Eng. Felinto Junior Da Costa
Londrina, 14 de janeiro de 2023.

Contents

	7
1 Um pouco da história	9
1.1 Primeiros levantamentos, estudos e publicações & Demografia e aritmética política	9
1.2 Visualização de dados & Estudos e primeiras publicações	14
1.3 Nomes notáveis	21
1.4 Revista Biometrika	22
1.5 Eugenia	22
2 Conceitos gerais	27
2.1 Estatística descritiva	27
2.2 Estatística inferencial	28
2.3 Produção de conhecimento	28
2.4 População (universo) & amostra	30
2.5 Parâmetros e estatísticas	31
2.6 Tipos de variáveis	31

Chapter 1

Um pouco da história

1.1 Primeiros levantamentos, estudos e publicações & Demografia e aritmética política

1086

O *Domesday Book* ([link](#)) foi encomendado em dezembro de 1085 por Guilherme, o Conquistador (*King William I*), que invadiu a Inglaterra em 1066.

O primeiro esboço foi concluído em agosto de 1086 e continha registros de 13.418 assentamentos nos condados ingleses ao sul dos rios Ribble e Tees (a fronteira com a Escócia) com informações sobre terras, proprietários, uso da terra, empregados e animais cujo propósito básico era fundamentar a taxação.

1602

O dramaturgo inglês William Shakespeare usou a palavra **statists** (estadistas e, portanto, num sentido não relacionado com números ou matemática) no diálogo da Cena II de Hamlet ([link](#)).

“Hamlet: Cercado assim por tantas vilanias, mesmo antes de eu poder dizer o prólogo, representava o cérebro. Sentei-me e escrevi com capricho nova carta. Já pensei, como os nossos estadistas, que é feio escrever bem, tendo insistido, até, em desaprendê-lo; mas, nessa hora muito bom me foi isso. Quererias saber qual o conteúdo da mensagem? [...]”

1603

O negociante inglês John Graunt (1620-1674) substituiu a crença pela evidência em *Natural and Political Observations Mentioned in a Following Index and Made upon the Bills of Mortality* (Observações naturais e políticas feitas sobre as notas de mortalidade).

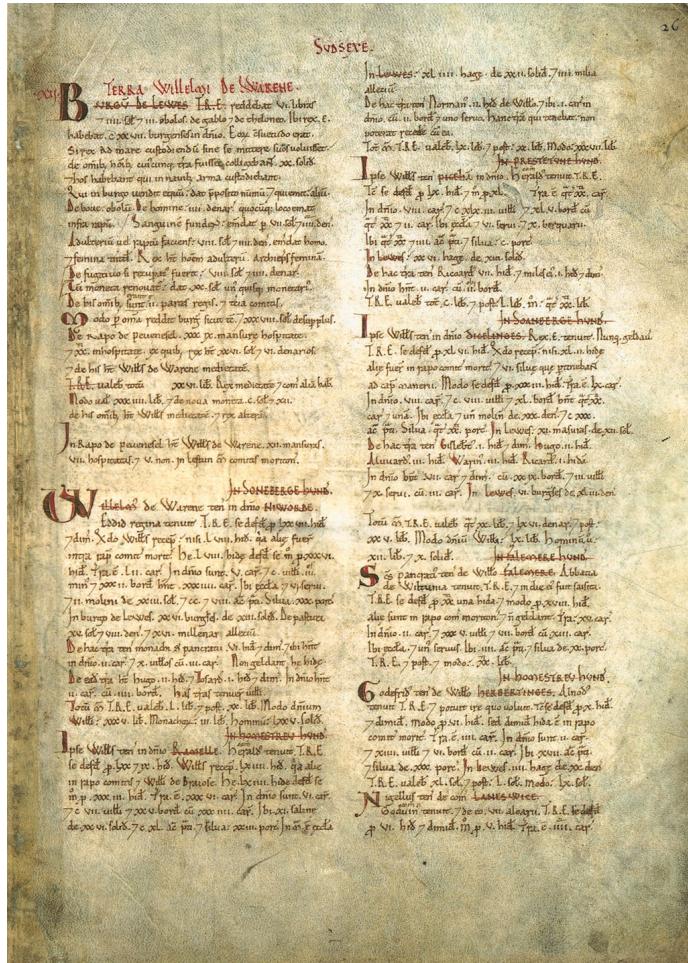


Figure 1.1: Domesday Book

1.1. PRIMEIROS LEVANTAMENTOS, ESTUDOS E PUBLICAÇÕES & DEMOGRAFIA E ARITMÉTICA POLÍTICA

Nesse trabalho, realizado com dados coletados das paróquias de Londres entre 1604 e 1660, Graunt tirou as seguintes conclusões: que havia maior nascimento de crianças do sexo masculino, mas havia distribuição aproximadamente igual de ambos os sexos na população geral; alta mortalidade nos primeiros anos de vida; maior mortalidade nas zonas urbanas em relação às zonas rurais.

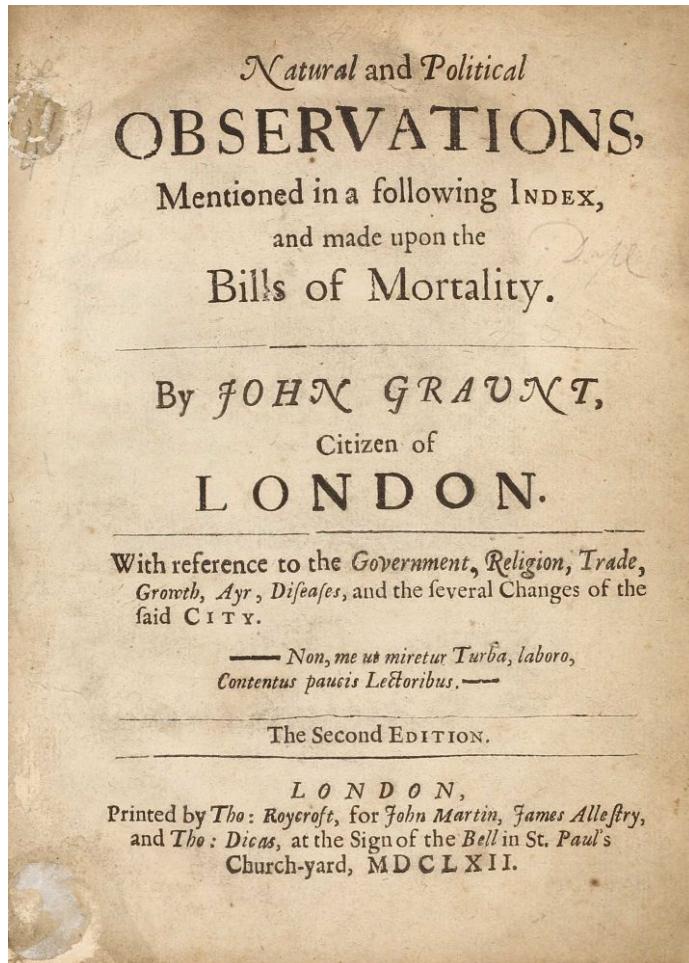


Figure 1.2: Natural and Political Observations Mentioned in a Following Index and Made upon the Bills of Mortality (ed. de 1662)

1660

Herman Conring (1606-1681), professor de filosofia, medicina e política da Universidade de Helmstadt (atual Alemanha), criou um curso de Ciência política em 1660, que descrevia e examinava as questões fundamentais do Estado. Nele a **estatística** passou a ser considerada como uma disciplina autônoma que tinha

por objetivo a descrição das coisas do Estado.

1687

Em 1687 o economista e filósofo inglês William Petty (1623-1687) publicou *Five Essays on Political Arithmetic* (Cinco ensaios sobre aritmética política), sugerindo ao governo inglês a criação de um departamento para registro de **estatísticas** vitais.

O Capitão John Graunt e William Petty instituiram na Inglaterra um novo ramo de estudos denominado de *Political arithmetic* (Aritmética política)

1693

O matemático e astrônomo inglês Edmond Halley (1656-1742) construiu em 1693, baseado em dados coletados na cidade (à época) alemã de Bresláu, uma *Life Table* (Tábua de sobrevivência), um estudo que analisa as probabilidades de sobrevivência e morte em relação à idade.

1749

Com um sentido não relacionado com números ou matemática, a palavra **estatística** parece ter sido proposta pela primeira vez no século XVII, pelo historiador e professor alemão (à época Transilvânia) Martin Schmeitzel (1679-1747) da Universidade de Jena e, posteriormente adotada por seu aluno, (igualmente) historiador e jurista Gottfried Achenwall (1719-1772) em 1749, em *Abriß der neuen Staatswissenschaft der vornehmen Europäischen Reiche und Republiken* (Esboço da nova ciência política dos nobres impérios europeus e repúblicas).

1771

William Hooper usou a palavra **estatística** em sua tradução de *The Elements of Universal Erudition* (Elementos da Erudição Universal) escrita por Jacob Friedrich Freiherr von Bielfeld (1717-1770).

Nesse livro, a **estatística** foi definida como a ciência que nos ensina o arranjo político de todos os estados modernos do mundo conhecido (mais uma vez num sentido não associado a números ou matemática).

1790

O jurista e político escocês John Sinclair propôs que se realizasse uma detalhada pesquisa em 938 paróquias para elucidar a história natural e política de seu país (*Statistics Accounts*). Essa pesquisa fazia parte de um projeto muito maior: *The Pyramid of Statistical Enquiry* (A Pirâmide da Pesquisa Estatística).

1854

O médico inglês (considerado por alguns como o “pai” da epidemiologia moderna) John Snow (1813-1858) estudou a dispersão espacial dos casos de cólera em Londres e concluiu que sua causa residia na contaminação da água consumida (poço localizado na *Broad Street*, no distrito do *Soho*): *Report to the Cholera Outbreak in the Parish of St. James, Westminster during the Autumn of*

1.1. PRIMEIROS LEVANTAMENTOS, ESTUDOS E PUBLICAÇÕES & DEMOGRAFIA E ARITMÉTICA POLÍTICA

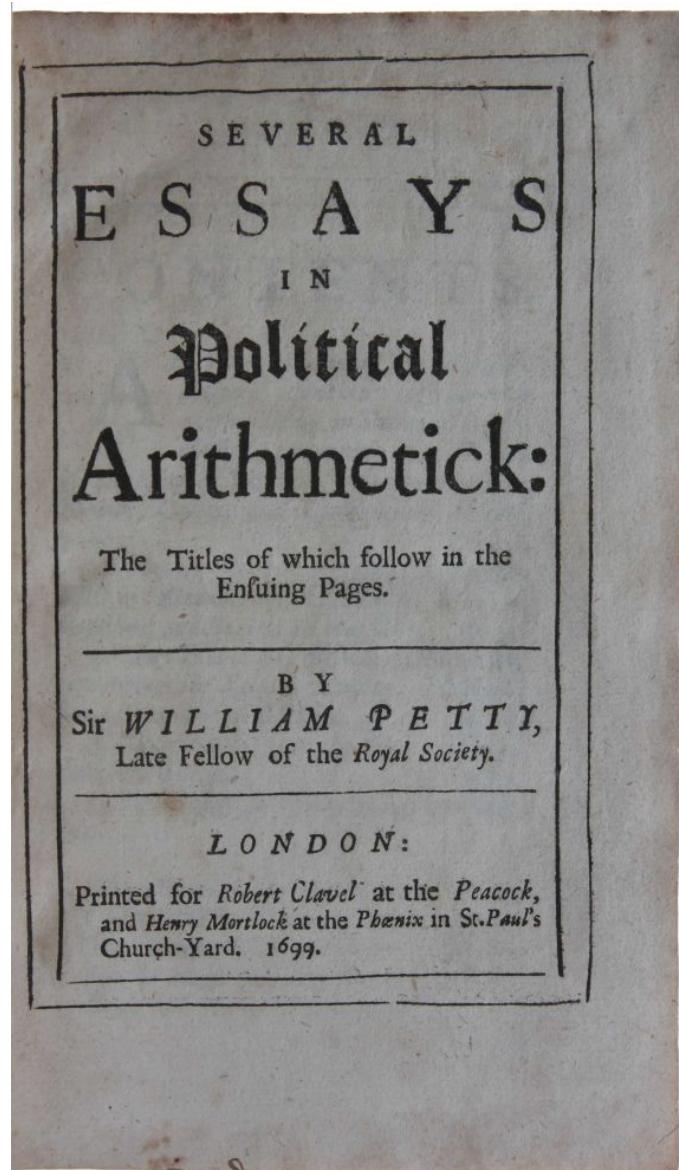


Figure 1.3: Several Essays in Political Arithmetick (ed. de 1699)

Age.	Per- sons.	Age.	Per- sons.										
Curt.		Curt.											
1	1000	8	680	15	628	22	585	29	539	36	481	7	5547
2	855	9	670	16	622	23	579	30	531	37	472	14	4584
3	798	10	661	17	616	24	573	31	523	38	463	21	4270
4	760	11	653	18	610	25	567	32	515	39	454	28	3504
5	732	12	646	19	604	26	560	33	507	40	445	35	3604
6	710	13	640	20	598	27	553	34	499	41	430	42	3198
7	692	14	634	21	592	28	546	35	490	42	427	49	2709
												56	2194
Age.	Per- sons.	Age.	Per- sons.										
Curt.		Curt.		Curt.		Curt.		Curt.		Curt.		Curt.	
43	419	50	346	57	272	64	202	71	131	78	58	77	692
44	407	51	335	58	262	65	192	72	120	79	49	84	253
45	397	52	324	59	252	65	182	73	109	80	41	100	107
46	387	53	313	60	242	67	172	74	98	81	34		
47	377	54	302	61	232	68	162	75	88	82	28		34000
48	367	55	292	62	222	69	152	76	78	83	23		
49	357	56	282	63	212	70	142	77	68	84	20		
												Sum Total.	

Figure 1.4: Halley's life table (1693)

1854 (Relatório sobre o surto de cólera na paróquia de St. James, Westminster durante o outono de 1854).

1.2 Visualização de dados & Estudos e primeiras publicações

1765

O teólogo e filósofo inglês Joseph Priestley (1733-1804) introduziu como inovação os primeiros gráficos com linha temporal, em que barras individuais eram usadas para visualizar o tempo de vida de uma pessoa e o todo pode ser usado para comparar a expectativa de vida de várias pessoas.

1786

O engenheiro e economista escocês William Playfair (1759-1823) é considerado comumente como fundador dos métodos gráficos para apresentação de estatísticas. Playfair concebeu vários tipos de diagramas para visualização de dados:

- em 1786, o gráfico de barras; e,
- em 1801, o gráfico de setores.

1856

A enfermeira inglesa Florence Nightingale (1820-1910) conduziu um trabalho pioneiro ao chegar no hospital militar britânico na Turquia em 1856, estabelecendo uma ordem e um método muito necessários aos registros médicos estatísticos e que indicaram serem as precárias práticas sanitárias o culpado da alta mortalidade ([link](#)).



Figure 1.5: Abriß der neuen Staatswissenschaft der vornehmen Europäischen Reiche und Republiken (1749)

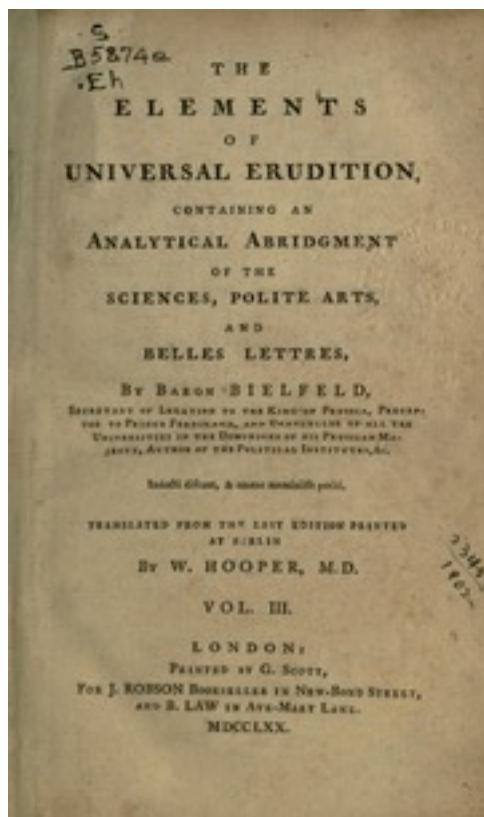


Figure 1.6: The Elements of Universal Erudition (1771)

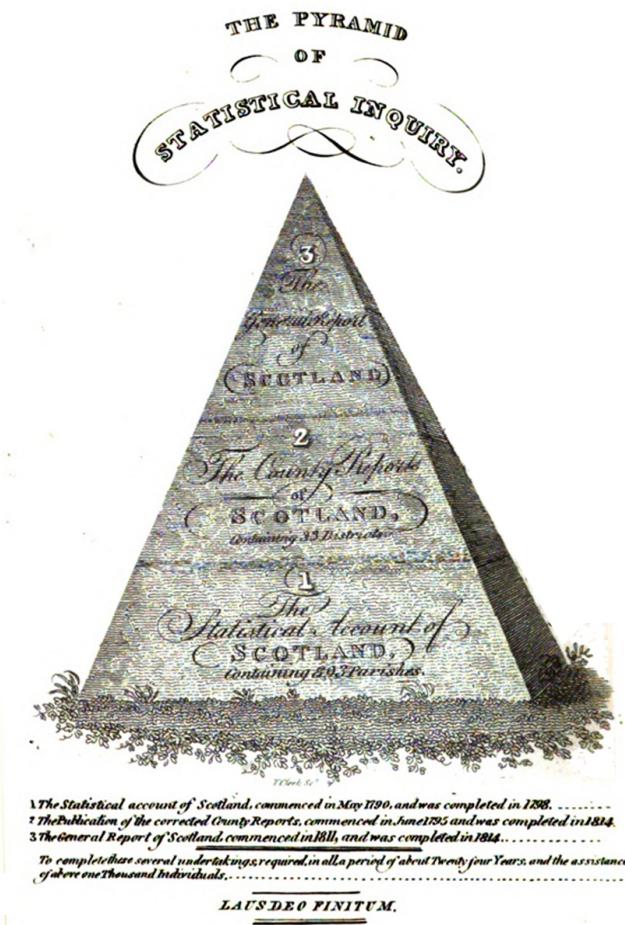


Figure 1.7: The Pyramid of Statistical Enquiry (1814)

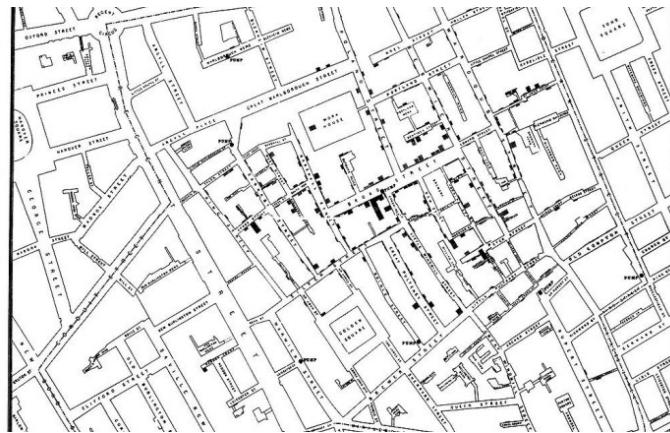


Figure 1.8: Mapa dos casos de cólera (1854)

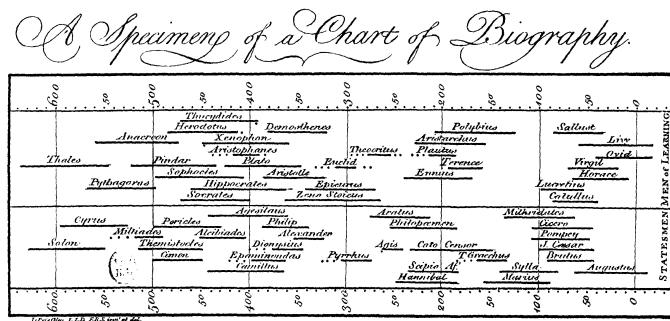


Figure 1.9: Expectativa de vida de diversas pessoas (1765)

1.2. VISUALIZAÇÃO DE DADOS & ESTUDOS E PRIMEIRAS PUBLICAÇÕES 19

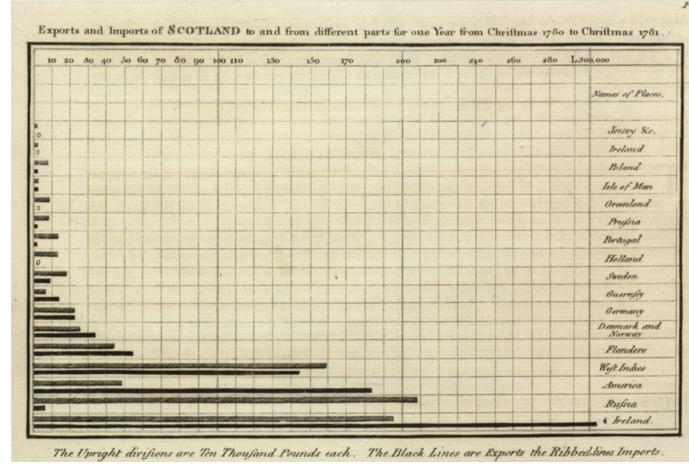


Figure 1.10: Commercial and Political Atlas (Atlas Comercial e Político de 1786): cada barra representa as exportações e importações da Escócia para 17 países em 1781

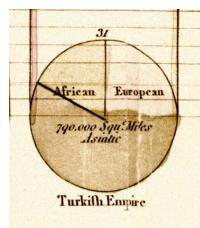


Figure 1.11: Statistical Breviary (Breviário Estatístico de 1801): proporção da extensão do Império Turco em diferentes regiões do mundo: Ásia, Europa e África, antes de 1789

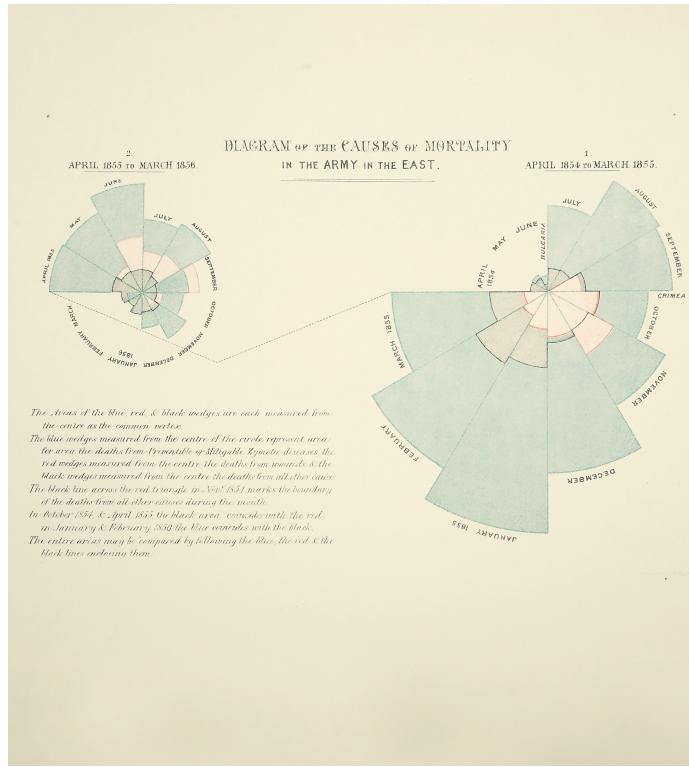


Figure 1.12: Esse diagrama (coxcomb) feito durante a Guerra da Crimeia foi dividido igualmente em 12 setores, representando os meses do ano, com a área sombreada do setor de cada mês proporcional à taxa de mortalidade naquele mês. Seu sombreamento com código de cores indicava a causa da morte em cada área do diagrama

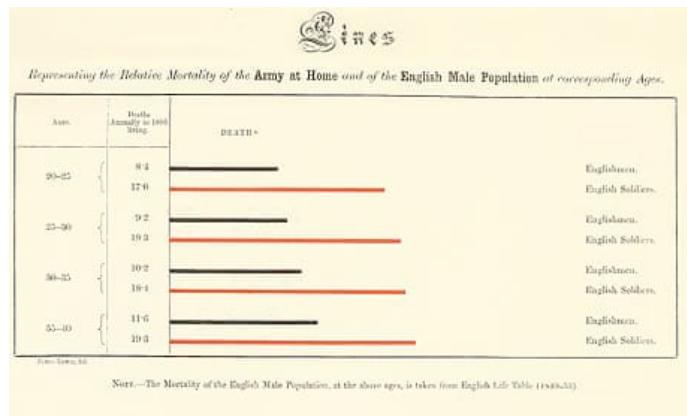


Figure 1.13: Gráfico de barras de Florence Nightingale mostrando as diferenças de mortalidade entre soldados britânicos e a população masculina inglesa geral (civis)

1.3 Nomes notáveis

Karl Pearson (1857-1936) é amplamente considerado o fundador da disciplina moderna de **estatística**, e também é famoso como um filósofo da ciência, como escritor sobre o darwinismo social e como um dos principais impulsionadores para instalar a eugenica como a ciência social chave. Uma breve biografia de cada um dos pesquisadores a seguir relacionados pode ser obtida em: (link).

- Niccolò Fontana Tartaglia (Veneza à época, hoje Itália: 1499-1557)
- Girolamo Cardano (Pávia à época, hoje Itália: 1501-1576)
- Galileu Galilei (Florença à época, hoje Itália: 1564-1642)
- Pierre de Fermat (França: 1607-1665)
- Blaise Pascal (França: 1623-1662)
- Jakob Bernoulli (Suíça: 1655-1705)
- Abraham de Moivre (França: 1667-1754)
- Thomas Bayes (Inglaterra: 1702-1761)
- Pierre-Simon Laplace (França: 1749-1827)
- Johann Carl Friedrich Gauss (Alemanha: 1777-1856)
- Lambert Adolphe Jacques Quêtelet (França à época, hoje Bélgica: 1796-1874)
- Pafnuti Lvovitch Chebyshev (Rússia: 1821-1894)
- Francis Galton (Inglaterra: 1822-1911)
- Wilhelm Lexis (Alemanha: 1837-1914)
- Thorvald Nicolai Thiele (Dinamarca: 1838-1910)
- Friedrich Robert Helmert (Saxônia: 1843-1917)
- Francis Ysidro Edgeworth (Inglaterra: 1845-1926)
- James Douglas Hamilton Dickson (Escócia: 1849-1931)
- Andrei Andreyevich Markov (Rússia: 1856-1922)

- Aleksandr Mikhailovich Lyapunov (Rússia: 1857-1918)
- Walter Frank Raphael Weldon (Inglaterra: 1860-1906)
- Karl Pearson (Inglaterra: 1857-1936)
- William Seally Gosset (Inglaterra: 1876-1937)
- Ronald Aylmer Fisher (Inglaterra: 1890-1962)
- Andrei Nikolaevich Kolmogorov (Rússia: 1903-1987)

1.4 Revista Biometrika

“Pretende-se que a *Biometrika* sirva como um meio não apenas de coletar ou publicar, sob um título, dados biológicos de um tipo não coletados sistematicamente ou publicados em outro lugar em qualquer outro periódico, mas também de disseminar um conhecimento de tal teoria estatística para o seu tratamento científico[...]”

Em outubro de 1901 foi fundada a *Biometrika, the Journal for the Statistical Study of Biological Problems* (*Biometrika*, o Jornal para o Estudo Estatístico de Problemas Biológicos) com o propósito de promover a análise estatística de fenômenos biológicos, isto é, a matematização da biologia.

Os fundadores da *Biometrika* foram Sir Francis Galton (primo de Charles Darwin), Walter Frank Raphael Weldon e Karl Pearson. A maior parte do trabalho foi feita por Pearson e Weldon, este último focando na edição do conteúdo (ou seja, o aspecto biológico) e o primeiro nos detalhes, incluindo correções de prova. Galton e o eugenista americano Charles Davenport atuaram, respectivamente, como consultor e editor.

Alguns dos tópicos abordados na revista incluem criminologia, botânica, zoologia, epidemiologia e outros aspectos da saúde humana. Na década de 1930, o caráter da *Biometrika* mudou, e “representou a vanguarda internacional da pesquisa em métodos estatísticos e sua aplicação na ciência e tecnologia”, ao invés de focar a hereditariedade.

Sir Francis Galton, que serviu como editor da primeira edição (1901), escreveu a Introdução, que incluiu uma declaração de propósito para a revista ([link](#)).

1.5 Eugenia

Em 16 de maio de 1883 Sir Francis Galton cunhou o termo “eugenia”, posteriormente descrevendo-o como “o estudo das agências sob controle social que podem melhorar ou reparar as qualidades raciais das gerações futuras, seja fisicamente ou mentalmente”.

Galton detalha o conceito em seu livro *Inquiries into Human Faculty and its Development*, e recomenda que indivíduos de famílias altamente classificadas em seu sistema de mérito sejam encorajados a se casar cedo e receber incentivos

para ter filhos. Ele também condenou os casamentos tardios dentro desse mesmo grupo como “disgênicos” ou desvantajosos para a espécie humana.

A palavra “eugenia” foi extraída da palavra grega *eu*, que significa bem, e *genos*, que significa prole. Juntos, significa bem-nascido.

Este livro caiu em domínio público e pode ser lido na íntegra online. A caracterização original de eugenio de Galton pode ser encontrada na página 17 desta edição de domínio público (Parte 1 do pdf):

“uma breve palavra para expressar a ciência de melhorar o rebanho, que não está de modo algum confinado a questões de acasalamento criterioso, mas que, especialmente no caso do homem, toma conhecimento de todas as influências que tendem, mesmo que em grau remoto, a dar ao raças ou linhagens de sangue mais adequadas uma melhor chance de prevalecer rapidamente sobre os menos adequados do que teriam de outra forma [...]”(Galton, 1883, p.17)

Há poucos anos alguns grupos sociais viram no trabalho e opiniões de Fisher endossos ao colonialismo, à supremacia branca e à eugenia.

Outros grupos, todavia, afirmam que Fisher não era racista e eugenista, embora ele achasse que havia diferenças comportamentais e de inteligência entre os grupos humanos.

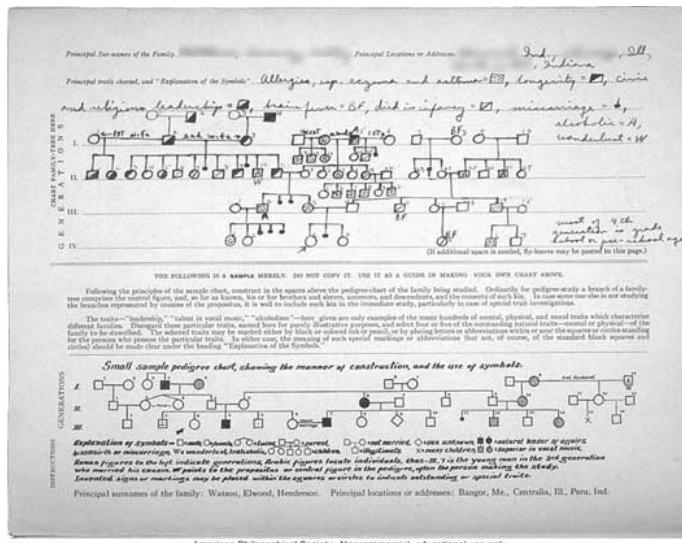


Figure 1.14: Gráfico de linhagens para alergias

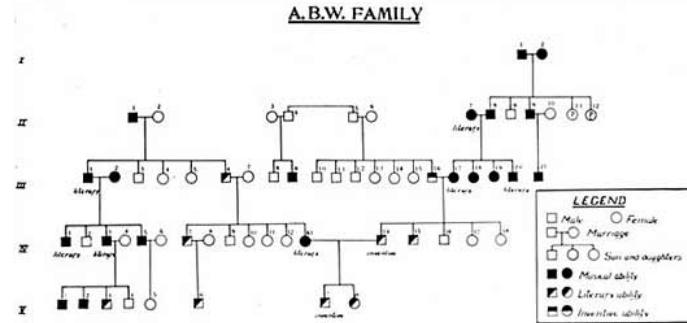


Figure 1.15: Gráfico de linhagens para aptidão musical

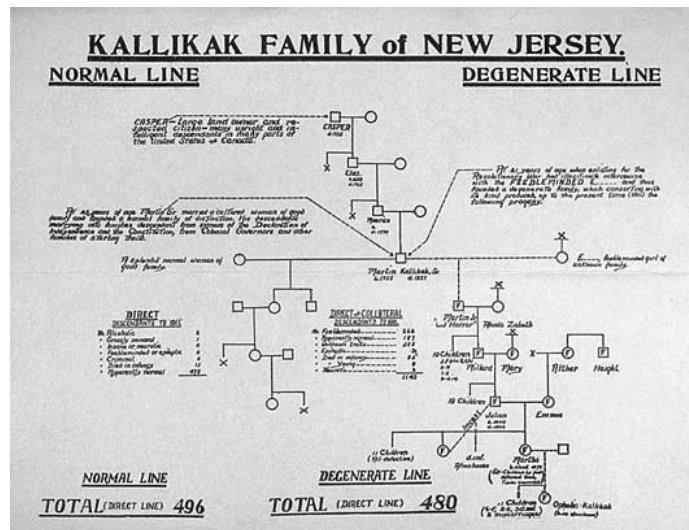


Figure 1.16: Linhas "normais" e "degeneradas" da família Kallikak (New Jersey)

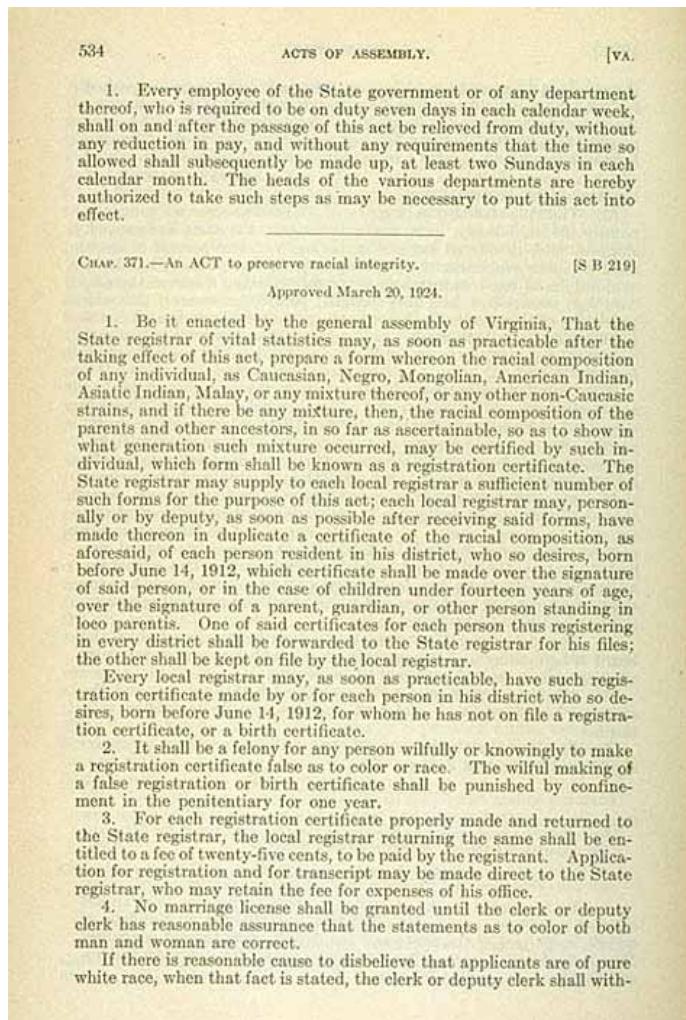


Figure 1.17: Lei da Inegridade Racia (Virginia, EUA, 1924)



Figure 1.18: Licença para casamento

Chapter 2

Conceitos gerais

“Estatística é um conjunto de métodos que se destina a possibilitar a tomada de decisões, face às incertezas[...]''

De modo geral, a estatística pode ser dividida em três grandes áreas:

- descritiva;
- probabilidade; e,
- inferencial.

2.1 Estatística descritiva

Nos primeiros trabalhos estatísticos, os dados coletados eram inicialmente apresentados na forma de tabelas e gráficos.

A **estatística descritiva** se ocupa de tudo o que seja relacionado a dados: coleta, processamento, descrição (seja na forma tabular ou gráfica) e sínteses numéricas (de locação, de dispersão, de repartição) sem inferir coisa alguma além da informação trazida pelos dados. Vem experimentando crescente uso em todas as áreas científicas e desenvolvimento:

- crescente uso de uma abordagem quantitativa em todas as ciências;
- disponibilidade de recursos computacionais;
- quantidade de dados coletados.

A palavra **estatística** pode assumir diferentes significados:

- no singular: **estatística**
 - refere-se à ciência que comprehende métodos que são usados na coleta, análise, interpretação e apresentação de dados quantitativos ou qualitativos (numéricos ou não); e,

- denota uma medida ou fórmula específica (tais como uma média, um intervalo de valores, uma taxa de crescimento, um índice).
- no plural: **estatísticas**
 - refere-se a dados coletados de maneira sistemática com um propósito específico definido em qualquer campo de estudo (nesse sentido, as *estatísticas* também podem ser consideradas como agregados de fatos expressos em forma numérica).

2.2 Estatística inferencial

A **estatística inferencial** tem o objetivo de estabelecer níveis de confiança da tomada de decisão de associar uma estimativa amostral a um parâmetro populacional. Divide-se em estimação e testes de significância.

“Dedução e indução são procedimentos racionais que nos levam do já conhecido ao ainda não conhecido; isto é, permitem que adquiramos conhecimentos novos graças a conhecimentos já adquiridos.[...]"

Dedução.

Na dedução parte-se de uma verdade já conhecida para demonstrar que ela se aplica a todos os casos particulares iguais. Vai do geral ao particular.

Indução.

Na indução parte-se de alguns casos particulares iguais ou semelhantes para se estipular uma **lei geral**. Vai do particular ao geral.

Na dedução, dado **X**, infiro (concluo) **a, b, c, d**.

Na indução, dados **a, b, c, d**, infiro (concluo) **X**.

Exemplo: testes de aceleração (0-60 mph) feitos com 6 carros importados em 1999 resultaram nas seguintes medidas: 12,9 s; 16,50 s; 11,30 s; 15,20 s; 18,20 s e 17,70 s. Um estudo descritivo poderia afirmar que:

- metade dos dados coletados acelera de 0-60 mph em menos de 16,00 s; e
- a aceleração média de 0-60 mph é de 15,30 s.

Mas, a partir dessa amostra concluir que a aceleração média de **todos** os carros importados em 1999 seja de 15,30 s; ou, que **metade** dos carros importados em 1999 acelerem de 0-60 mph em menos de 16,00 s são afirmações que pertencem à **inferência estatística**.

2.3 Produção de conhecimento

Na expansão de qualquer área do conhecimento propomos hipóteses que serão avaliadas mediante a coleta de dados que, depois de analisados, revelarão infor-

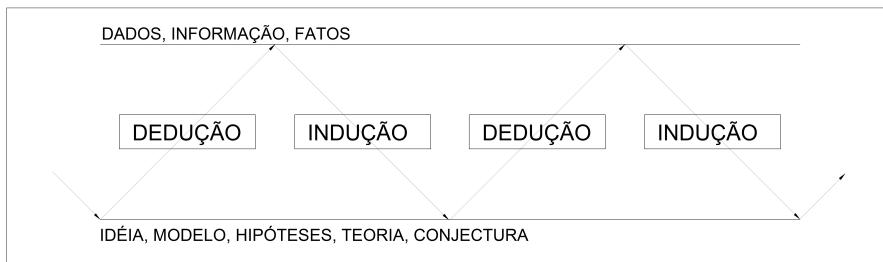


Figure 2.1: Fluxograma elementar de um processo de aprendizagem

mações que, eventualmente, nos conduzirão ao afastamento da hipótese original e à proposição de outras, num processo contínuo como, por exemplo:

(A)

- Hipótese (ideia, teoria, conjectura): “Hoje será um dia como outro qualquer.”
- Dedução: “Meu carro estará estacionado na garagem, no local de costume.”
- Dados (informação, fatos): “Meu carro não está lá!”
- Inferência: “Alguém deve tê-lo levado.”

(B)

- Hipótese (ideia, teoria, conjectura): “Meu carro foi roubado!”
- Dedução: “Meu carro não estará no local de costume.”
- Dados (informação, fatos): “Meu carro está lá!”
- Inferência: “Alguém deve tê-lo levado e devolvido.”

(C)

- Hipótese (ideia, teoria, conjectura): “Um ladrão pegou e trouxe de volta.”
- Dedução: “Meu carro foi arrombado.”
- Dados (informação, fatos): “Meu carro está intacto e o alarme está desligado.”
- Inferência: “Alguém que tenha as chaves deve tê-lo levado.”

(D)

- Hipótese (ideia, teoria, conjectura): “Minha esposa usou meu carro.”
- Dedução: “Ela provavelmente deixou um bilhete.”
- Dados (informação, fatos): “Sim, aqui está o bilhete.”
- Inferência: “Minha hipótese estava correta.”

Uma investigação científica deve envolver, em linhas gerais:

- observação dos fatos;
- descrição das características essenciais, segundo o que se obteve através da observação;
- explicação dessas características descritivas;
- previsão; e,
- decisão pertinente à investigação.

O planejamento de uma pesquisa deve envolver, em linhas gerais:

- definição do *universo*: é necessário delimitar claramente, no tempo e espaço, o âmbito do inquérito, definindo, em termos precisos, o *universo* a ser trabalhado;
- exame das informações disponíveis: deve-se reunir todo o material existente: mapas, artigos, livros, relatórios relativos a levantamentos semelhantes;
- tipos de levantamentos: completo ou amostral;
- prazo;
- custo;
- precisão.

2.4 População (universo) & amostra



Figure 2.2: Universo e amostra

Quase que, invariavelmente, em todo ramo de conhecimento, o pesquisador esbarra em uma série de limitações das mais variadas ordens (econômica, técnica, ética, geográfica, temporal,...) que impossibilitam o estudo dos dados e informações associados a todos os casos existentes (**população ou universo**).

Por essa razão, através de um procedimento estatístico denominado de

amostragem, estuda-se uma população (universo) a partir de uma amostra. Amostra é, portanto, um subconjunto finito e representativo da população (universo), extraído de modo sistemático (planejado).

2.5 Parâmetros e estatísticas

É comum a adoção de letras gregas para as características descritivas que se referirem à população (universo) e letras do alfabeto latino para aquelas relativas à amostra extraída:

Característica estudada	Notação populacional	Notação amostral
Número de elementos	N	n
Média	μ	\bar{x}
Variância	σ^2	s^2
Desvio padrão	σ	s
Proporção	Π	p

A α Alpha	B β Beta	$\Gamma\gamma$ Gamma	$\Delta\delta$ Delta	E ε Epsilon	Z ζ Zeta	H η Eta	$\Theta\theta$ Theta
I ι Iota	K κ Kappa	$\Lambda\lambda$ Lambda	M μ Mu	N ν Nu	$\Xi\xi$ Xi	O \circ Omicron	$\Pi\pi$ Pi
P ρ Rho	$\Sigma\sigma\varsigma$ Sigma	T τ Tau	Y υ Upsilon	$\Phi\phi$ Phi	X χ Chi	$\Psi\psi$ Psi	$\Omega\omega$ Omega

Figure 2.3: Alfabeto grego

2.6 Tipos de variáveis

Variáveis quantitativas

- contínuas: são os dados com maior potencial de produzir informação significativa dentre todos: comprimentos, áreas, pesos, densidades; e,
- discretas: são dados com um pouco menos de informação que os de natureza contínua mas possuem mais informação que dados qualitativos: número de andares de um prédio, de degraus de uma escada, número de filhos de um casal.

Variáveis qualitativas

- ordinais: apresentam um pouco mais de informação que os dados qualitativos puramente nominais na medida que suas classes podem ser interpretadas como possuindo um ordenamento inerente: padrão construtivo (baixo, médio, alto), classe econômica de rendimento (baixa, média, alta), nível de escolaridade (fundamental, médio e superior); e,
- nominais: são dados a menor quantidade de informação: sexo, cor, códigos postais de cidades;

Codificação de variáveis qualitativas

- binárias: pela associação de valores numéricos: 0 ou 1 a uma variável qualitativa nominal que se apresente com apenas dois aspectos: sim ou não, ausência ou presença. Pela composição de mais variáveis binárias pode-se codificar variáveis que possuam um número maior de classes; e,
- *proxy*: pela associação de valores numéricos contínuos que guardam “correlação” com as classes da variável qualitativa nominal.

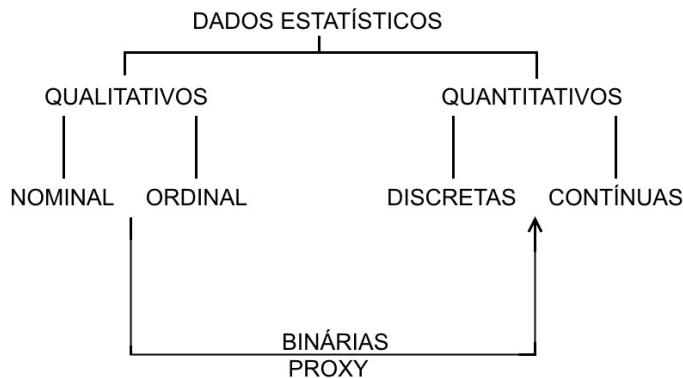


Figure 2.4: Tipos e codificações de variáveis