

História da Estatística

Gauss M. Cordeiro (UFRPE)

email: gauss@deinfo.ufrpe.br

1. O que é Estatística?
2. Origens dos Conceitos Estatísticos
3. Conceitos Básicos da Inferência Estatística
4. Áreas de Aplicação da Estatística
5. Avanços Recentes da Estatística
6. Estatística e sua História

O que é Estatística?

Antigamente: palavra latina STATUS (Estado)

Hoje: Estatística = Ciência + Tecnologia + Arte

Ciência: tem identidade própria com um grande número de técnicas deduzidas de princípios básicos

Tecnologia: sua metodologia pode ser implantada num sistema operacional para manter um nível desejado de performance

Arte: depende da razão indutiva e não está livre de controvérsias

Origens dos Conceitos Estatísticos

Problema Real

Análise de dados de astronomia.

Qual a distância de Berlin a Koln?

Os planetas descrevem órbitas distribuídas aleatoriamente?

Qual a população da França?

Hereditariedade Genética

Como melhorar a qualidade da cerveja?

Melhoramento de técnicas agrícolas.

Conceito Estatístico

Mínimos Quadrados

Testes de Hipóteses

Estimadores da razão

Regressão e Correlação

Teste t

Planejamento de Experimentos

Origens dos Conceitos Estatísticos

Problema Real

As variedades de batatas e fertilizantes interagem?

Explicando ciclos estranhos nos preços de trigo.

Inspeção de Munição

Jogos nos Cassinos de

Monte Carlo

Necessidade de realizar

inúmeros testes na Indústria

Farmacológica

Eficiência das Indústrias do

Japão após II Guerra Mundial

Conceito

Análise de Variância

Modelos Paramétricos

de Séries Temporais

Testes Sequenciais

Simulação Estocástica

Testes Não-Paramétricos

Técnicas de Controle de

Qualidade

Conceitos Básicos da Inferência Estatística

Teoria da Regressão	1885	Francis Galton
Teste Qui-quadrado	1900	Karl Pearson
Teste t	1908	W. S. Gosset
Estimação	1921	Sir Ronald Fisher
ANOVA	1923	Sir Ronald Fisher
Planej. de Experimentos	1926	Sir Ronald Fisher
Testes de Hipóteses	1933	Neyman & Pearson
Testes Não-Paramétricos	1937	E. J. G. Pitman
Planos Amostrais	1945	P. C. Mahalanobis
Teoria da Decisão	1950	A. Wald
Modelos de Sobrevivência	1972	Sir David R. Cox

Atuação Principal da Estatística

Bioestatística - Aplicações à Medicina

Biometria - Aspectos estatísticos e matemáticos da Biologia

Demografia - Ciência da população humana e sua evolução no tempo

Econometria - Estudo de problemas econômicos combinando métodos estatísticos e matemáticos com teoria econômica

Epidemiologia - campo da medicina que trata do estudo de epidemias

Sociometria - Estudo de problemas sociais através da Estatística

Áreas de Aplicação da Estatística

- Medicina (Diagnóstico, Prognóstico, Ensaios Clínicos)
- Genética e Epidemiologia
- Agricultura (Experimentação Agrícola)
- Indústria e Negócio (Controle de Qualidade, Previsão de Demanda, Gerenciamento Eficiente, Mercado e Finanças)
- Governo (Disseminação da Informação, Políticas de Decisão, Serviços Públicos)
- Economia (Técnicas Econométricas e Séries Temporais)
- Pesquisa (Artes, Arqueologia, Ciências Agrárias, Ciências Exatas, Ciências Sociais, Literatura, Meio Ambiente, Mercado, Petróleo)
- Direito (Evidência estatística, teste de DNA, investigação criminal)

Avanços Recentes

- Bootstrap
- Data Mining e Redes Neurais (Campo Multidisciplinar da Estatística que envolve: Estatísticas, Validação Cruzada, Sistemas Especialistas de Dados e Técnicas de Visualização)
- Modelos de Regressão (Modelos Lineares Generalizados, Modelos Aditivos Generalizados, Modelos de Dispersão)
- Métodos MCMC
- Sistemas Especialistas
- Geoprocessamento e Reconhecimento de Padrões

Produção Mundial entre 1985-1996

País	Páginas	Artigos
Estados Unidos	109338	4061
Reino Unido	12597	538
Canadá	12407	516
Austrália	7872	323
Alemanha	6782	306
França	3647	129
Japão	2865	163
Holanda	2864	116
Índia	2559	151
Israel	2097	83
(Brasil - 23a posição)	744	24

Histórico da Estatística (5000 - 310 AC)

5000 AC - Registros egípcios de presos de guerra

3000 AC - Jogos de Dados

2000 AC - Censo Chinês

1500 AC - Dados de mortos em guerras no Velho Testamento

1100 AC - Registros de dados em livros da Dinastia Chinesa

585 AC - Thales de Mileto usa a geometria dedutiva

540 AC - Pitágoras (Aritmética e Geometria)

430 AC - Philolaus obtém dados de Astronomia e Hippocrates estuda doenças a partir da coleta de dados

400 AC - Estabelecido o Censo Romano

310 AC - Descrição detalhada como os dados deveriam ser coletados em livros de Constantinopla

Histórico da Estatística (300AC-470)

300 AC - Elementos de Euclides

180 AC - Origem de Dados Circulares (Hypsicles)

140 AC - Surge a Trigonometria com Hipparchus

100 AC - Horácio usa um ábaco de fichas como
instrumento de “cálculo portátil”

120 - Menelaus apresenta tabelas estatísticas cruzadas

250 - Estudos Avançados na Aritmética por Diophantus

300 - Desenvolvimento da álgebra

400 - Desenvolvimento da teoria dos números

470 - Valor de pi por Tsu Chung-Chi

Histórico da Estatística (620-840)

- 620 - Surge em Constantinopla um Primeiro Bureau de Estatística que existe até 680
- 695 - Primeira utilização da média ponderada pelos árabes na contagem de moedas
- 775 - Trabalhos estatísticos hindus são traduzidos para o árabe
- 826 - Os árabes usam cálculos estatísticos na tomada de Creta
- 830 - Al-Khwarizmi desenvolve a álgebra
- 840 - O astrônomo persa Yahyâ Abî Mansûr apresenta tabelas astronômicas amplas

Histórico da Estatística (1202-1620)

- 1202 - Sequência de Fibonacci
- 1303 - Origem dos Números Combinatórios (Shihchieh Chu)
- 1405 - O persa Ghiyat Kâshî realiza os primeiros cálculos de probabilidade com a fórmula do binômio
- 1447 - Surgem as primeiras tabelas de mortalidade construídas pelos sábios do Islã
- 1530 - Lotto de Firenze – Primeira Loteria Pública
- 1550 - Número Combinatório (Cardano)
- 1572 - Origem dos números complexos (Bombelli)
- 1591 - Solução de uma equação cúbica (Viète)
- 1593 - Fórmula de Viète para π
- 1614 - Napier cria os logaritmos
- 1620 - Descartes descobre a Geometria Analítica

Histórico da Estatística (1629-1684)

- 1629 - Método de Máximo e Mínimo e Teoria dos Números (Pierre de Fermat)
- 1654 - Pierre de Fermat e Blaise Pascal estabelecem os Princípios do Cálculo das Probabilidades
- 1656 - Huygens publica o primeiro tratado de Probabilidade
- 1660 - Fundação da Royal Society of London
- 1662 - Primeiros estudos demográficos (Graunt)
- 1665 - Expansão do Binômio de Newton
- 1670 - Fórmula de Interpolação de Gregory-Newton e Último Teorema de Fermat
- 1673 - $\pi/4 = 1 - 1/3 + 1/5 - 1/7 + \dots$ (Leibniz)
- 1679 - Distribuição de Pascal, Tratado do Triângulo Aritmético e conceito do Valor Esperado (Pascal)
- 1684 - Leibniz desenvolve o Cálculo Diferencial e Integral

Histórico da Estatística (1687-1738)

1687 - Principia Matemática de Newton

1693 - Edmund Halley publica tabelas de mortalidade e cria os fundamentos da Atuária

1702 - Logaritmos Complexos (Bernoulli)

1707 - Fórmula de D'Moivre

1713 - Distribuição Binomial (Bernoulli)

1715 - Teorema de Taylor

1718 - D'Moivre publica Doutrina de Chances

1727 - Número "e" de Euler

1730 - Distribuição Normal (D'Moivre) e Fórmula de Stirling

1733 - Teorema Central do Limite (D'Moivre)

1738 - Medição do Risco (Daniel Bernoulli)

Histórico da Estatística (1748-1765)

- 1748 - Fórmula do Produto (função zeta em termos de potências dos primos) e Identidade de Euler
- 1749 - Método Minimax (Euler)
- 1750 - Teorema dos Números Pentagonais (Euler)
- 1753 - Solução Geral da Equação da Onda (Bernoulli)
- 1756 - Distribuições Discretas uniforme e do triângulo isósceles (Simpson)
- 1763 - Inferência Bayesiana (Thomas Bayes)
- 1764 - Probabilidade Condicional e Teorema de Bayes
- 1765 - Distribuição Contínua Semi-Circular (Lambert)

Histórico da Estatística (1774-1799)

- 1774 - Teoria da Estimação e Distribuição Exponencial Dupla (Laplace)
- 1775 - William Morgan se torna o Primeiro Atuário
- 1776 - Distribuições Contínuas Uniforme e Parabólica (Lagrange)
- 1777 - Primeiro exemplo de uso da verossimilhança na estimação de parâmetro (Daniel Bernoulli)
- 1797 - Funções Analíticas (Lagrange) e o verbete Statistics aparece na Enciclopédia Britânica
- 1799 - Mecânica Celeste (Laplace)

Histórico da Estatística (1800-1821)

1800 - A França estabelece o seu Bureau de Estatística

1804 - Análise de dados da órbita do Halley (Bessel)

1805 - Método dos Mínimos Quadrados (Legendre)

1810 - Teorema Central do Limite (Laplace)

1812 - Théorie Analytique des Probabilités (Laplace)

1820 - Várias sociedades de Estatística são criadas

1821 - Demonstração do que se chama hoje Teorema de Gauss-Markov (Gauss)

Histórico da Estatística (1822-1853)

1822 - Séries de Fourier

1826 - Princípio da Dualidade (Poncelet)

1827 - Movimento Browniano

1831 - Teoria Geral das Equações de Galois

1834 - Primeiro Computador Analítico (Charles Babbage) e
Fundação do Journal of the Royal Statistical Society-B

1835 - Lei dos Grandes Números (Poisson)

1836 - Distribuição Gama

1837 - Distribuição de Poisson

1839 - Fundação da American Statistical Association (ASA)

1846 - Uso de Quantis (Quetelet)

1853 - Distribuição de Cauchy e Primeira Conferência Internacional de
Estatística em Bruxellas (Quetelet)

Histórico da Estatística (1856-1873)

- 1856 - Artur Cayley cria o Cálculo Matricial
- 1859 - Função Zeta com argumentos complexos (Riemann)
- 1860 - Polinômios de Chebyshev-Hermite
- 1863 - Distribuição Qui-Quadrado (Abbé)
- 1864 - Distribuição de Hermite
- 1867 - Desigualdade de Chebyshev
- 1869 - Genialidade Hereditária (Galton)
- 1871 - Notas em Hospitais (Florence Nightingale)
- 1873 - Determinação experimental de pi e prova por Hermite que “e” é transcendental

Histórico da Estatística (1876-1896)

- 1876 - Primeiro uso de um Método do tipo Monte Carlo (Forest)
- 1879 - Super-Dispersão de Dados (Lexis)
- 1882 - Prova que pi é transcendental por Lindemann
- 1885 - Fundação do ISI
- 1887 - Teoria da Regressão (Galton) e Índice de Marshall
- 1889 - Princípios da Lógica Indutiva (Venn)
- 1892 - Coeficiente de Correlação (Edgeworth)
- 1894 - Método dos Momentos e uso pela primeira vez dos termos momento e desvio padrão (Karl Pearson)
- 1895 - Sistema de Distribuições e Coef. de Variação (Pearson)
- 1896 - Métodos de Captura-Recaptura (Petersen)

Histórico da Estatística (1897-1908)

- 1897 - Distribuição de Pareto
- 1900 - Teste Qui-Quadrado (K. Pearson), Cadeias de Markov e Coeficiente de Yule
- 1901 - Fundação da Biometrika (Pearson, Weldon e Galton)
- 1903 - Semi-Invariantes ou Cumulantes (Thiele)
- 1904 - Análise Fatorial, Expansão de Edgeworth, Coeficiente de Contingência (Karl Pearson) e Coeficiente de Spearman
- 1905 - Curva de Lorenz
- 1906 - Cálculo Funcional (Frechet)
- 1908 - Distribuição t de Student e distribuição amostral do coeficiente de correlação (William Gosset) e Análise Fatorial (Spearman)

Pearson

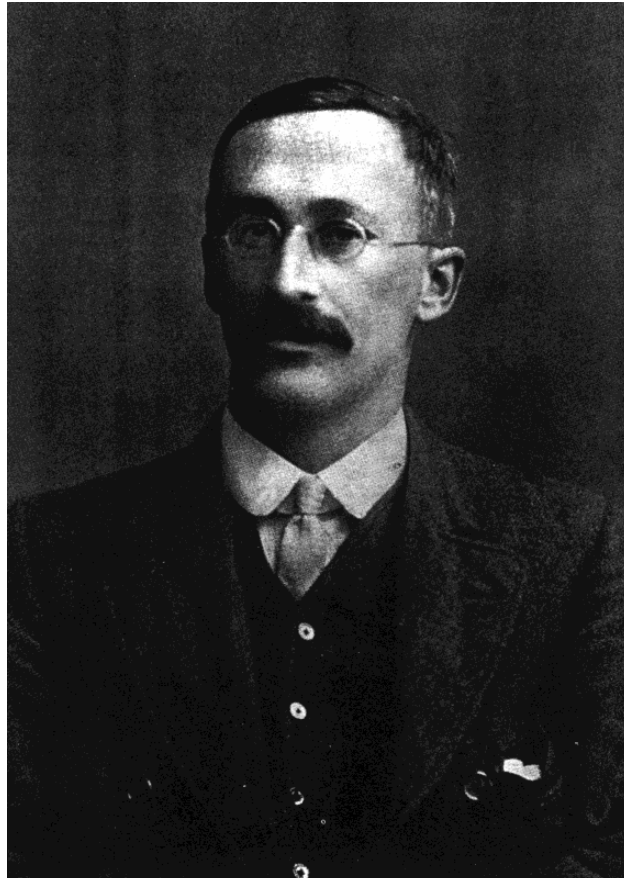


Markov



А. А. Марков (1886).

Gosset



'Student' in 1908

Histórico da Estatística (1912-1923)

- 1912 - Método de Máxima Verossimilhança (Ronald Fisher) e Índice de Gini
- 1913 - Primeiro Teorema da Teoria dos Jogos (E. Zermelo), Bureau de Estatísticas do Trabalho dos EUA e Desigualdade de Markov
- 1914 - Método do Mínimo Qui-Quadrado (Yule)
- 1917 - Fórmula do Atraso de Erlang
- 1918 - Definição formal de Variância em um artigo de Genética (Fisher) e Distribuição de von Mises
- 1921 - Informação e Suficiência (Fisher), Probabilidade Lógica (Keynes)
- 1922 - Definição de Verossimilhança, Consistência e Eficiência (Fisher) e Prova Rigorosa do Teorema Central do Limite (Lindeberg)
- 1923 - Tabela ANOVA (Fisher) e Processo de Wiener

Histórico da Estatística (1924-1930)

1924 - Desigualdade de Bernstein

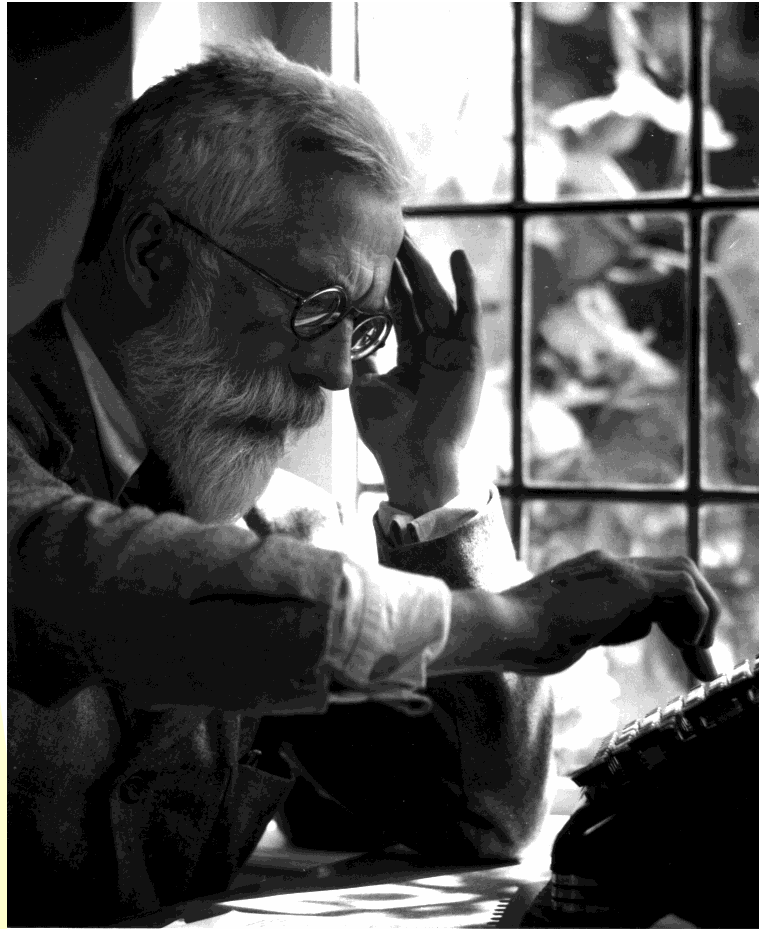
1925 - Método score para parâmetros, definição de p-valor e “Statistical Methods for Research Workers” (Fisher)

1926 - Planejamento de Experimentos (Fisher) e Conceito de Hipótese Alternativa (Gosset)

1928 - Distribuições Não-Centrais (Fisher), Intervalos de Confiança, Razão de Verossimilhanças e Poder dos Testes (Neyman & Pearson) e Distribuição de Wishart

1930 - Controle de Qualidade nas indústrias, Inferência Fiducial (Fisher), Distância de Mahalanobis, Tempo Médio de Espera na Fila M/G/1 (Pollaczek) e Fundação da Econometrica

Fisher



Histórico da Estatística (1931-1935)

- 1931 - Definição de Espaço Amostral (von Mises) e Cartas de Controle de Qualidade (Shewhart)
- 1932 - Distribuição de Gumbel
- 1933 - Lema de Neyman & Pearson, Distância de Kolmogorov, Componentes Principais (Hotteling) e Fundamentos de Probabilidade (Kolmogorov)
- 1934 - Estatística Ancilar, Família Exponencial e Princípios da Verossimilhança (Fisher), Teorema de Cochran, Distribuição F (Snedecor) e Análise de Confluência (Frisch)
- 1935 - Curva de Mortalidade-Dosagem (Bliss), Formulação Matemática da Família Exponencial (Darmois) e Correlação Canônica (Hotteling)

Histórico da Estatística (1936-1944)

- 1936 - Verossimilhança Marginal e Condicional (Bartlett), Desigualdades de Bonferroni, Função Suporte (Jeffreys) e Estatística de Smirnov
- 1937 - Inferência Não-Paramétrica (Pitman), Teoria das Regiões de Confiança (Neyman), Expansão de Cornish-Fisher, Correção de Bartlett e Ensaios Clínicos Aleatorizados (Hill)
- 1938 - Distribuição Assintótica da Razão de Verossimilhança (Wilks)
- 1939 - Distribuição de Weibull e início dos Métodos Bayesianos (Jeffreys)
- 1940 - Invenção do Computador Eletrônico
- 1944 - Surgem as Técnicas de Monte Carlo

Histórico da Estatística (1945-1947)

- 1945 - Planos Amostrais (Mahalanobis), Desigualdade de Cramér-Rao, Teorema de Rao-Blackwell, Testes Sequenciais (Wald) e Teste de Wilcoxon
- 1946 - Condições de Regularidade do EMV (Cramér), Distrib. Log-Gama (Bartlett e Kendall), Distribuição a priori de Jeffreys e Estatísticas U e V (Halmos)
- 1947 - Distribuição Normal Inversa e Métodos Sequenciais (Wald), Teste de Mann-Whitney, Estatística score (Rao), Modelos Exponenciais de Dispersão (Tweedie) e Método Simplex (Dantzig)

Histórico da Estatística (1949-1953)

- 1949 - Método de Linearização, Eficiência em Grandes Amostras (Neyman) e Teste de Aditividade de Tukey
- 1950 - Teoria Estatística de Decisão (Wald) e Teorema de Lehmann-Scheffé
- 1951 - Modelo de Regressão Heterocedástico, Testes Não-Paramétricos de Lehmann e Primeiro Computador Comercial instalado no Bureau de Censo dos EUA
- 1952 - Estatística de Anderson-Darling e Teste de Kruskal-Wallis
- 1953 - Método de amostragem que envolvia Cadeias de Markov (Metropolis et. al.) e Inferência Robusta (Box)

Histórico da Estatística (1954-1960)

- 1954 - Aproximações Ponto de Sela (Daniels) e Carta CUSUM (Page)
- 1955 - Completude e Regiões Similares (Lehmann e Scheffé)
- 1956 - Método Jackknife (Quenouille)
- 1957 - Programação Dinâmica (Bellman)
- 1958 - Estimador de Kaplan-Meier
- 1959 - Estudo retrospectivo de doenças (Mantel & Haenszel)
- 1960 - Inferência em modelos de espaço de estados (Kalman) e Comparação Estocástica (Bahadur)

Histórico da Estatística (1961-1970)

1961 - Famílias Separadas de Hipóteses (Cox) e Filtro de Kalman

1962 - Princípios de Inferência (Birnbaum)

1964 - Modelos de Box e Cox e Estimação Robusta (Huber)

1967 - Caso Multivariado da Expansão de Edgeworth (Chambers)

1968 - Inferência Estrutural (Fraser)

1970 - Modelos ARMA (Box & Jenkins) e Modelos Log-Lineares, Generalização do Método de Metropolis (Hastings) e Regressão Rígida (Hoerl e Kennard)

Histórico da Estatística (1972-1977)

- 1972 - Modelo de Riscos Proporcionais (Cox) e os Modelos Lineares Generalizados (Nelder & Wedderburn)
- 1973 - Critério da Informação de Akaike
- 1974 - Quase-Verossimilhança (Wedderburn)
- 1975 - Curvatura Estatística (Efron), Verossimilhança Parcial (Cox) e Teoria da Catástrofe (René Thom)
- 1976 - Enfoque Bayesiano em Modelos de Espaço de Estados (Harrison e Stevens)
- 1977 - Análise Exploratória de Dados e Distribuições g e h (Tukey), Algoritmo EM (Dempster, Laird e Rubin) e EMV em pequenas amostras (Bowman & Shenton)

Histórico da Estatística (1979-1985)

- 1979 - Método Bootstrap (Efron) e Verossimilhança Preditiva (Mathiasen)
- 1980 - Aproximações Ponto de Sela para Soma Estocástica (Lugannani e Rice) e Teste de White
- 1981 - Surgimento do IBM-PC e Estimadores M (Huber)
- 1982 - Modelos ARCH (Engle) e Redes Neurais (Hopfield)
- 1983 - Fórmula p^* e verossimilhança perfilada (Barndorff-Nielsen)
- 1984 - Amostrador de Gibbs (Geman & Geman)
- 1985 - Modelos para Análise de Dados Longitudinais, Inferência Pivotal (Barnard) e Cálculo de PI com 17 milhões de dígitos

Histórico da Estatística (1986-1990)

- 1986 - GEE (Equações de Estimação Generalizadas) de Liang e Zeger, Modelos Aditivos Generalizados (Hastie & Tibshirani) e Modelos GARCH (Bollerslev)
- 1987 - Definição de Yoke (Barndorff-Nielsen) e Modelos de Dispersão (Jorgensen)
- 1988 - KDD (“Knowledge Discovery in Databases”) e Modelos Dinâmicos (West e Harrison)
- 1990 - Métodos MCMC no contexto Bayesiano (Gelfand e Smith), Mineração de Dados (“Data Mining”), Momentos L (Hosking), Teoria da Perturbação Estocástica (Stewart)

Histórico da Estatística (1991-2002)

- 1991 - Computação Neural (Hertz, Krogh e Palmer) e Mínimos Quadrados Total (Van Huffel e Vandewalle)
- 1995 - Modelos Multiníveis (Goldstein)
- 1996 - Profundidade da Regressão (Rousseeuw e Hubert)
- 1997 - Modelos Fatoriais
- 2001 - 100 anos da Biometrika
- 2002 – Cálculo de PI com 51 bilhões de dígitos

Mercado de Trabalho

Pesquisas de Opinião

Pesquisas Eleitorais

Perfis de Consumidores

Indústrias

Controle de Qualidade

Avaliação de Desempenho

Instituições Públicas:

IBGE

Ministérios e Secretarias de Estado

DATAPREV

...

Instituições de Ensino e Pesquisa - Universidades

Mercado de Trabalho

Instituições Financeiras (bancos, seguradoras, ...)

Cálculo de apólices de seguro

Construção de indicadores econômicos

Carteiras de investimentos

Construção de plano de pensão

Instituições da Área Médica (hospitais, clínicas,...)

Eficácia de novos tratamentos

Controle de infecção hospitalar

Controle de Qualidade

Contexto: uma linha de produção industrial.

Problemas:

(1) Como o fabricante pode garantir que o processo de produção está sob controle?

(2) Como o fabricante pode garantir que as especificações do produto estão de acordo com as indicadas na embalagem?