

Elementos de Estatística

Lupércio F. Bessegato & Marcel T. Vieira

UFJF – Departamento de Estatística
2013



Apresentação

- Lupércio França Bessegato
lupercio.bessegato@ufjf.edu.br
Departamento de Estatística
- Marcel de Toledo Vieira
marcel.vieira@ufjf.edu.br
Departamento de Estatística

Ementa

- Conceitos básicos:
 - √ Introdução à metodologia estatística
 - √ Variáveis e níveis de mensuração
 - √ Variabilidade
- Descrição Tabular e Gráfica:
 - √ Tabelas
 - √ Gráfico de setores, de barras, histogramas, box-plot, diagrama de dispersão, de Pareto, etc.

- Medidas resumo de conjunto de dados:
 - √ Média, mediana, moda
 - √ Variância e desvio-padrão
 - √ Quartis e percentis. Distância interquartílica
 - √ Assimetria e curtose
- Séries Estatísticas:
 - √ Série temporal, espacial
 - √ Números índices. Índices agregativos

- Modelos Probabilísticos:
 - √ Probabilidade: introdução, eventos, Bayes
 - √ Modelo gaussiano
- Noções de Inferência:
 - √ Amostragem:
 - noções e técnicas de amostragem, survey
 - √ Teste de hipótese de trabalho
 - √ Teste de Qui-quadrado

Bibliografia

- Wild, C.J. e Seber, G.A.F. (LTC) *Encontros com o Acaso: um Primeiro Curso de Análise de Dados e Inferência*
- Moore, D. S. e McCabe, G. P. (LTC) *Introdução à Prática da Estatística*
- Agresti, A. e Finlay, B. (Penso) *Métodos Estatísticos para Ciências Sociais*

Questionário

Roteiro

1. Conceitos Básicos	}	Análise Exploratória
2. Descrição Tabular e Gráfica		
3. Medidas-resumo		
4. Séries Estatísticas		
5. Regressão e Correlação		
6. Amostragem	}	Inferência
7. Modelos Probabilísticos		
8. Noções de Inferência		

Conceitos Básicos

O que é *Estatística*?

- Segundo *Magalhães e Lima (2005)*, *Estatística* é um conjunto de técnicas que permite, de forma sistemática, organizar, descrever, analisar e interpretar dados oriundos de estudos ou experimentos, realizados em qualquer área do conhecimento.

Importância da *Estatística*

- Utilização rotineira de conceitos estatísticos
 - √ Comportamento dados climáticos
 - √ Distribuição espacial de dados epidemiológicos
 - √ Dados sócio-econômicos, etc.
- Uso métodos estatísticos na literatura
- Pesquisa de fenômenos geográficos
- Aplicação na investigação CIENTÍFICA

Estatística

- Conjunto de técnicas para:
 - √ Organizar dados
 - √ Descrever dados
 - √ Analisar dados
 - √ Interpretar dados oriundos:
 - estudos
 - Experimentos
- Aplicadas em qualquer área do conhecimento

Algumas Áreas de Aplicação

- Demografia e Economia:
 - √ Índices econômicos, índices demográficos, aprimoramento de processos de gerenciamento
- Saúde:
 - √ Previsão de surtos epidêmicos, eficiência de tratamentos, estudo de incidência de doenças
- Agricultura:
 - √ Aprimoramento de produtos agrícolas

- Engenharia:
 - √ Desenvolvimento de equipamentos, controle de tráfego, controle de qualidade, estudos hidrográficos
- Setor Público:
 - √ IBGE: produção estatísticas para apoiar estudos e planejamentos de políticas públicas
- Política:
 - √ Pesquisas eleitorais, pesquisas de opinião pública

- Marketing:
 - √ Lançamento de novos produtos
- Educação:
 - √ Avaliação de sistemas educacionais

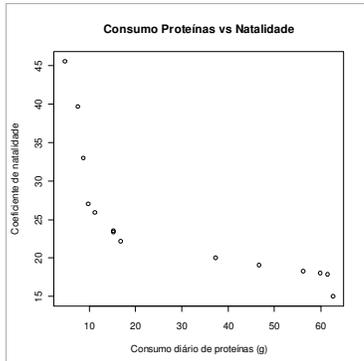
Geografia – Exemplos

- Como se dá a variação espacial dos níveis de chumbo no sangue de crianças?
 - √ Níveis estão espalhados aleatoriamente ou existe um padrão geográfico discernível?
 - √ Como os padrões estão relacionados às características da residência dos moradores?

- É possível descrever a difusão geográfica da democracia que ocorreu no período de pós IIª Guerra como um processo contínuo ao longo do tempo ou ela ocorreu em ondas?
- Quais os efeitos do aquecimento global na distribuição geográfica das espécies?
 - √ Que mudanças ocorrerão nos tipos e na distribuição espacial das espécies de árvores em áreas específicas?

- Quais os efeitos de diferentes estratégias de marketing no desempenho do produto?
 - √ As estratégias de marketing de massa são eficazes, apesar de localizadas mais distantes de seus mercados?

Gráficos – Exemplos



- Consumo de gordura e câncer de mama (1979)

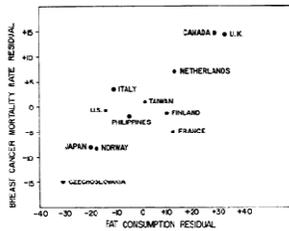
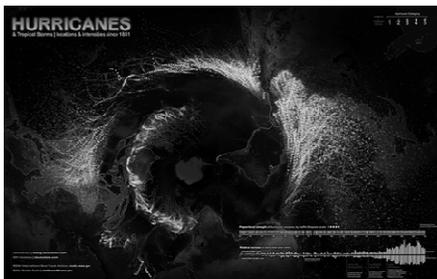


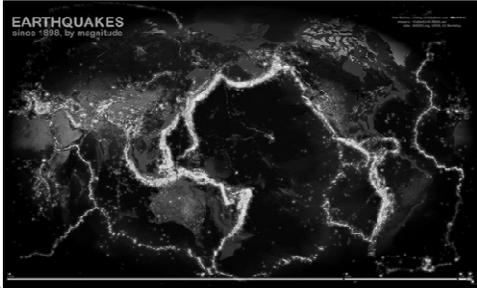
FIG.—Relationship between breast-cancer mortality-rate residuals (after fitting to height, weight and age at menarche) and total fat-consumption residuals (after fitting to height, weight and age at menarche). Mortality-rate residuals are per 100,000 women aged 35–64, and total fat-consumption residuals are in grams per day.

✓ Um século e meio de furacões, tornados, ciclones e tufões



- ✓ Centrado no Polo Sul
- ✓ Dados entre 1851 e 2010
- ✓ Trajetórias de furacões e tempestades tropicais

√ Todos os terremotos de pouco mais de um século



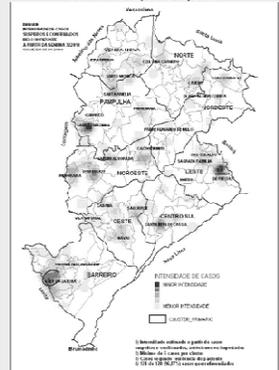
√ Sismos registrados desde 1898

√ Cada ponto luminoso é o epicentro.

- Quanto mais brilhante, mais tremores.
- Magnitudes separadas por cores (de 4 a 8 na escala Richter.)

Representação Espaci

Mapa 1. Intensidade dos casos confirmados de dengue, Belo Horizonte, SE 32-36/2011.



Fonte: Sistema Qualiquê e SISVA/FIOCRUZ/MS/ARCA/708
incluindo casos importados - 05/05/2011

Estatística – Áreas

- Estatística Descritiva:
 - √ Análise Exploratória de Dados
- Probabilidade
- Inferência Estatística

Definições Importantes

- Dados:
 - √ observações de um mesmo evento (acontecimento)
- População (ou Universo):
 - √ Conjunto global, sobre o qual se deseja chegar a conclusões
- Amostra:
 - √ sub-conjunto finito de uma população
 - √ Conjunto parcial sobre o qual são feitas observações.

Estatística Descritiva

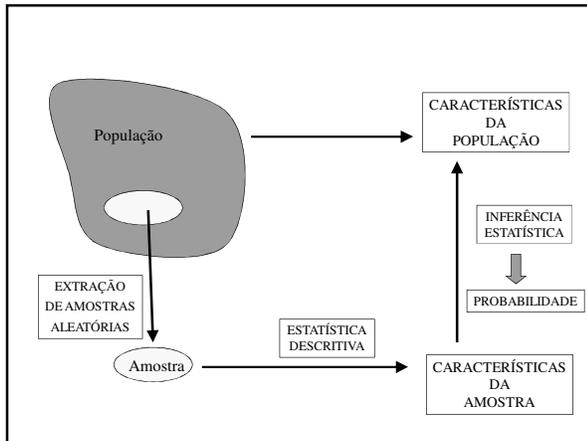
- Objetivo:
 - √ Estabelecer hipóteses a respeito de características de interesse
- Etapa inicial da análise:
 - √ Primeiro contato com os dados
- Técnicas para descrever e resumir os dados
 - √ Tabelas, gráficos e medidas-resumo

Probabilidade

- Teoria matemática que estuda a **incerteza**
 - √ Estudo de fenômenos aleatórios
- Fenômenos Aleatórios
 - √ Acontecimentos cujos resultados **não podem** ser previstos com certeza

Inferência Estatística

- Técnicas que possibilitam a extrapolação de conclusões
- Amostra → Universo
- Sub-conjunto → Conjunto-universo



Problema Fundamental da Estatística

A partir da observação de amostras, COMO podemos tirar CONCLUSÕES sobre a POPULAÇÃO ?

Dados Brutos

- Obtidos diretamente de pesquisa
 - √ Ainda sem qualquer processo de síntese ou análise
- Incluídos em tabelas
 - √ Porém, não incluídos em publicações

Variável

- Característica que pode variar no seu valor entre sujeitos (indivíduos) em uma amostra ou população
 - √ Sujeitos diferentes podem ter valores diferentes de uma variável

Escala de Mensuração

- Valores que a variável pode ter
- Exemplo:
 - √ Sexo: masculino e feminino
 - √ Número de filhos: 0, 1, 2, ...
- Os métodos estatísticos válidos para uma variável dependem de sua escala de mensuração

Variáveis

- Quantitativa:
 - √ Escala de mensuração tem valores numéricos
 - √ Os valores representam magnitudes diferentes da variável
- Qualitativas:
 - √ Escala de mensuração é um conjunto de categorias
 - √ Categorias distintas diferem em qualidade, não em magnitude numérica

- Escala intervalar (variáveis quantitativas):
 - √ Têm uma distância numérica específica (intervalo) entre cada par de níveis
- Escala nominal (variáveis qualitativas):
 - √ Escala não tem uma extremidade 'superior' ou 'inferior'
- Escala ordinal (variáveis qualitativas):
 - √ Escala categórica tendo uma ordem natural de valores

Variáveis Quantitativas

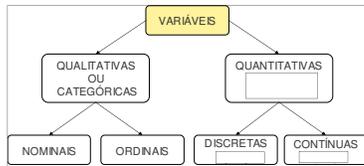
- Variável discreta:
 - √ seus valores possíveis formam um conjunto separado (enumerável) de números como 0, 1, 2, 3, ...
- Variável contínua:
 - √ pode ter um contínuo de valores reais possíveis

Comentários

- Aspectos quantitativos de dados ordinais
- Métodos estatísticos para variáveis discretas são usados principalmente para variáveis quantitativas que aceitam relativamente poucos valores
- Na prática, as variáveis quantitativas que podem aceitar muitos valores são tratadas como *contínuas*.

Variável

- Característica de interesse qualquer, cujo valor pode variar em cada “indivíduo”
- Classificação:



Variável Categórica Nominal

- Valores distribuídos em categorias mutuamente exclusivos
- Exemplos:
 - ✓ Sexo
 - ✓ Causa da morte
 - ✓ Região de procedência

Variável Categórica Ordinal

- Valores distribuídos em categorias mutuamente exclusivas que possuem uma ordenação natural
- Exemplos:
 - √ Grau de instrução
 - √ Classe sócio-econômica

Variáveis Quantitativas Discretas

- Dado um valor, é possível estabelecer seu sucessor
- Em geral, são resultados de contagens
- Exemplos:
 - Qte. de moradores em domicílio
 - Qte. de batimentos cardíacos

Variáveis Quantitativas Contínuas

- Podem assumir qualquer valor em um intervalo contínuo
- Em geral, são resultados de medições:
 - √ Temperatura
 - √ Peso
 - √ Altura

Variáveis

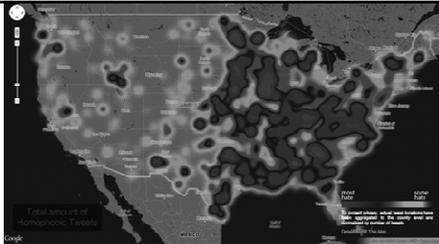
- Qualquer característica associada a um elemento pertencente a uma população ou uma amostra
- Classificação de variáveis:

Qualitativa { Nominal Sexo, cor dos olhos
Ordinal Classe social, grau de instrução

Quantitativa { Discreta Número de filhos, nº de carros
Contínua Peso, altura, salário

Atividade nº 1

ID	Idade	Sexo	Salário	Altura	Filhos	Forma	Idade	Sexo	Idade	Idade	Idade	Idade	
1	A	F	17	160	60,5	2	NAO	F	3	1	B	16	R
2	A	F	18	169	55,0	1	NAO	M	3	1	B	7	R
3	A	M	18	186	72,8	2	NAO	F	5	2	M	16	R
4	A	M	24	184	59,9	3	NAO	F	6	3	R	20	R
5	A	F	19	159	55,0	1	NAO	M	2	2	B	5	R
6	A	M	19	176	60,0	3	NAO	M	2	1	B	2	R
7	A	F	20	160	59,0	1	NAO	F	3	1	B	7	R
8	A	F	18	164	47,0	1	SIM	I	2	2	M	10	R
9	A	F	18	162	67,8	3	NAO	M	3	3	M	12	R
10	A	F	17	164	50,0	2	NAO	M	2	2	M	10	R
11	A	F	18	172	70,0	1	SIM	I	10	2	B	8	N
12	A	F	18	166	54,0	3	NAO	M	3	2	B	0	R
13	A	F	21	170	58,0	2	NAO	M	5	1	M	30	R
14	A	M	19	179	60,5	1	SIM	I	5	1	M	2	N
15	A	F	18	165	63,5	1	NAO	I	4	1	R	10	H
16	A	F	19	163	47,4	3	NAO	R	3	1	B	18	R
17	A	F	17	164	60,0	1	NAO	F	3	1	B	10	N
18	A	M	18	180	59,2	2	NAO	F	3	4	B	10	R
19	A	F	20	169	54,8	1	NAO	F	3	2	D	5	R
20	A	F	18	168	52,5	3	NAO	M	7	2	B	14	M
21	A	F	21	170	60,0	3	NAO	F	5	2	A	6	R
22	A	F	18	165	68,8	1	NAO	M	3	3	B	6	R
23	A	F	18	157	49,2	1	SIM	I	5	4	B	10	R
24	A	F	23	159	49,0	1	SIM	I	3	1	M	28	H
25	A	F	20	169	51,8	2	NAO	F	5	5	M	4	N
26	A	F	19	154	57,0	2	NAO	I	5	2	B	5	R
27	B	F	23	162	63,0	2	NAO	M	5	2	M	5	R
28	B	F	18	162	62,0	1	NAO	F	1	1	M	10	R
29	B	F	18	157	48,0	2	NAO	F	3	1	B	12	R



- ✓ Verificados tweets com termos contra raça, sexualidade e pessoas com deficiência
- ✓ 150.000 tweets foram agregados em nível de condado 'geotagged'
- ✓ Quantidades foram normalizadas pelo número de tweets em cada condado

Referências

Bibliografia

- Magalhães, M.N. e Lima, A.C.P.L. (Edusp)
Noções de Probabilidade e Estatística
- Wild, C.J. e Seber, G.A.F. (LTC)
Encontros com o Acaso: um Primeiro Curso de Análise de Dados e Inferência
- Agresti, A. e Agresti, B.F. (Dellen Pub.)
Statistical Methods for the Social Sciences
