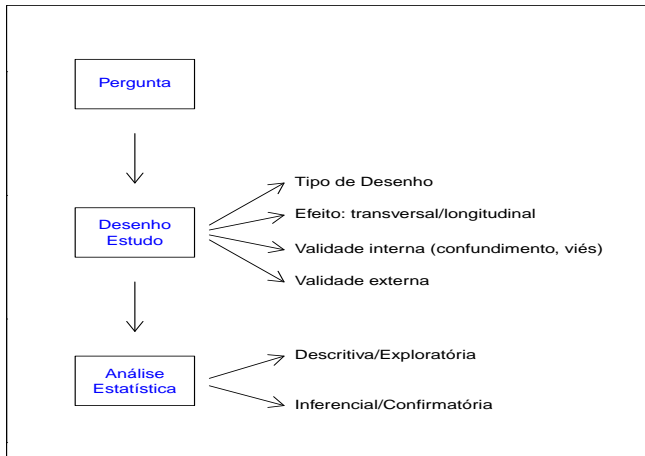


Métodos Estatísticos Avançados em Epidemiologia

Introdução à Disciplina

Enrico A. Colosimo/UFMG

Depto. Estatística - ICEx - UFMG



- 1 Pergunta de Interesse Clínico;
- 2 Desenho do Estudo;
- 3 Análise Estatística
 - Conhecer o Banco de Dados;
 - Análise Descritiva (cada variável separadamente)
 - Análise Univariada (ou bivariada) (resposta vs cada covariável)
 - **Análise Multivariada (tema principal da disciplina)**

Perguntas de Interesse/Curiosidade Clínica

1 Inferência

- Comparação de Grupos.
- Identificação de Fatores de Risco ou Prognóstico.

2 Estimação/Predição.

Desenho do Estudo

- 1 Tipo de Desenho de Estudo.
- 2 Transversal vs Longitudinal.
- 3 Viés (confundimento, seleção, etc).
- 4 Validação do Estudo.

Perguntas Relevantes

- Os grupos são comparáveis?
- As variáveis de confusão foram medidas/controladas?
- É possível alocar tratamento às unidades amostrais de forma aleatória?
- Os erros de medição podem ser medidos e controlados?
- As perdas (dados perdidos) podem viciar os resultados?
- Podemos estender os resultados para outros estudos?
- Existe efeito de calendário (ou de coorte)?

Tipos de Estudos Epidemiológicos

- 1 Estudos Descritivos: sem grupo de comparação
 - Ecológicos;
 - Transversais.
- 2 Estudos Analíticos: com grupo(s) de comparação
 - Observacionais;
 - Coorte (prospectivo ou histórico);
 - Caso-controle (retrospectivo);
 - Estudos Experimentais: Ensaio Clínico ou do tipo "cross-over".

A Prospective Study of Cigarette Smoking and the Incidence of Diabetes Mellitus among US Male Physicians

JoAnn E. Manson, MD, DrPH, Umed A. Ajani, MBBS, MPH, Simin Liu, MD, ScD,
David M. Nathan, MD, Charles H. Hennekens, MD, DrPH

PURPOSE: To determine the association between cigarette smoking and the incidence of type 2 diabetes mellitus.

SUBJECTS AND METHODS: We studied 21,068 US male physicians aged 40 to 84 years in the Physicians' Health Study who were initially free of diagnosed diabetes mellitus, cardiovascular disease, and cancer. Information about cigarette smoking and other risk indicators was obtained at baseline. The primary outcome was reported diagnosis of type 2 diabetes mellitus.

RESULTS: During 255,830 person-years of follow-up, 770 new cases of type 2 diabetes mellitus were identified. Smokers had a dose-dependent increased risk of developing type 2 diabetes mellitus: compared with never smokers, the age-adjusted relative risk was 2.1 (95% confidence interval [CI]: 1.7 to 2.6) for

current smokers of ≥ 20 cigarettes per day, 1.4 (95% CI: 1.0 to 2.0) for current smokers of < 20 cigarettes per day, and 1.2 (95% CI: 1.0 to 1.4) for past smokers. After multivariate adjustment for body mass index, physical activity, and other risk factors, the relative risks were 1.7 (95% CI: 1.3 to 2.3) for current smokers of ≥ 20 cigarettes per day, 1.5 (95% CI: 1.0 to 2.2) for current smokers of < 20 cigarettes per day, and 1.1 (95% CI: 1.0 to 1.4) for past smokers. Total pack-years of cigarette smoking was also associated with the risk of type 2 diabetes mellitus (P for trend < 0.001).

CONCLUSIONS: These prospective data support the hypothesis that cigarette smoking is an independent and modifiable determinant of type 2 diabetes mellitus. *Am J Med.* 2000;109:538-542. ©2000 by Excerpta Medica, Inc.

Um Estudo sobre Hipertensão Arterial Sistêmica na Cidade de Campo Grande, MS

A Study on Systemic Arterial Hypertension in Campo Grande, MS, Brazil

Ana Rita Araújo de Souza, Anselmo Costa, Diogo Nakamura, Leandro Nascimento Mocheti, Paulo Roberto Stevanato Filho, Luiz Alberto Ovando

Universidade para o Desenvolvimento do Estado e da Região do Pantanal, Hospital Beneficente Santa Casa de Campo Grande, Sociedade Brasileira de Cardiologia – MS - Campo Grande, MS

Resumo

Objetivos: Detectar a real prevalência de hipertensão arterial sistêmica em Campo Grande, MS, e fatores frequentes.

Métodos: Estudo transversal com amostra randomizada da população adulta da cidade de Campo Grande, MS, num total de 892 pessoas. Foi aplicado questionário sobre idade, sexo, escolaridade, tabagismo, etilismo, aspectos sobre o tratamento. Foram colhidos dados antropométricos (peso e altura). Segundo a OMS, foi considerado peso normal: $IMC < 25 \text{ kg/m}^2$; sobrepeso: $25 \leq IMC < 30$; obeso: $IMC \geq 30$. Os critérios para hipertensão foram baseados no VII Joint, com valores de corte de Pressão Arterial de $140 \times 90 \text{ mmHg}$.

Resultados: A prevalência de hipertensão foi de 41,4%, variando conforme idade (até 29 anos: 11,8%; 30-39: 24,8%; 40-49: 43,3%; 50-59: 42,4%; 60-69: 48,6% e ≥ 70 : 62,3%). Houve maior prevalência nos homens (51,8%), enquanto nas mulheres foi de 33,1%. As pessoas com formação escolar de 1º grau primário tendem a apresentar maiores índices pressóricos. Nos indivíduos com sobrepeso e obesidade, observou-se maior prevalência de pressão elevada: IMC normal (27,9%), sobrepeso (45,6%) e obesidade (58,6%). A partir dos 60 anos existe um maior percentual de hipertensão sistólica isolada, representado por 16,4% (60-69 anos) e de 24,6% (> 70 anos). Etilismo diário ou semanal também está relacionado a maior incidência, respectivamente, de 63,2% e 47,2%. Apenas 59,7% eram sabidamente hipertensos. Das pessoas que apresentaram hipertensão, 57,3% fazem algum tratamento. Dos que fazem tratamento regularmente, 60,5% apresentaram hipertensão.

Conclusão: A prevalência de hipertensão foi de 41,4%, ultrapassando a média detectada em alguns trabalhos, alertando para piora epidemiológica e repercussões cardiovasculares, o que evidencia necessidade de maior investimento público no que tange ao esclarecimento e instrução desses grupos populacionais quanto à prevenção.

Palavras-chave: Hipertensão/prevalência, hipertensão/epidemiologia, fatores de risco.

Arq Bras Cardiol 2007; 88(4)

Framingham Risk Score Calculator for Coronary Heart Disease

This **Framingham risk score calculator** estimates the 10-year coronary heart disease risk of any person based on certain criteria like gender, age, cholesterol and systolic pressure. You can discover more about this heart disease scoring system and about all the cardiovascular risk factors involved below the form.

Gender:*

Select ▼

Age:*

Total cholesterol (mg/dL):*

HDL cholesterol (mg/dL):*

Under hypertension treatment?

Select ▼

Systolic blood pressure (mmHg):*

Smoker?

Select ▼

Calculate

The Patient Age is required!The Total cholesterol in mg/dL is required!The HDL cholesterol in mg/dL is required!The Systolic blood pressure in mmHg is required!

The 10-year cardiovascular risk for coronary heart disease (CHD) is 5%.

Disclaimer: This tool should NOT be considered as a substitute for any professional medical service, NOR as a substitute for clinical judgement.

1 Resposta/Desfecho

- Quantitativa/Contínua;
- Qualitativa/Categórica.

2 Covariáveis/Exposição

- Interesse: Exposição;
- Confundimento.

- 1 Desvio da verdade: defeito no delineamento ou na condução de um estudo.
- 2 Viés: Erro sistemático
 - no delineamento, condução ou análise de um estudo;
 - efeito: erro na estimativa da magnitude da associação entre covariáveis/exposição e o desfecho de interesse.

1 Confundimento.

2 Viés de Seleção

Alocação das unidades amostrais privilegia subgrupos com probabilidade diferenciada de apresentar a resposta.
Exemplo: Perda de acompanhamento em estudos longitudinais.

3 Viés de Informação

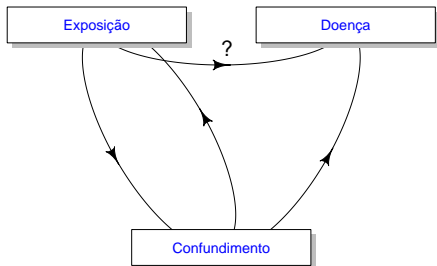
Erro sistemático na classificação de exposição ou da resposta.
Exemplo: testes diagnósticos.

4 Viés de Publicação

Exemplo: Revisão sistemática e meta-análise.

Confundimento

Definição: Um terceiro fator que está associado tanto com a exposição quanto com o desfecho, mas não se encontra no elo causal entre eles.



Seta unidirecional: indica associação causal e

Seta bidirecional: indica associação não causal.

Confundimento

- Definição: Um terceiro fator que está associado tanto com a exposição quanto com o desfecho, mas não se encontra no elo causal entre eles.
- Duas condições para ser fator de confundimento:
 - Ser associado com a exposição sem ser sua consequência.
 - Estar associado com o desfecho/doença independente da exposição.
- No elo causal: fator de mediação (usualmente, após a intervenção).

Confundimento: Exemplos

- Idade na associação entre câncer de pulmão e fumo.
- Fumo na associação entre câncer de estômago e café.
- (contra-exemplo: no elo causal?) Colesterol na associação entre infarto e dieta

Viés de Seleção (Estudos Longitudinais)

- O viés de seleção pode ocorrer devido a seguimento incompleto dos indivíduos do estudo (perdas de seguimento).
- O principal problema, mesmo para perdas pequenas, é a probabilidade de que elas possam ter ocorrido de forma seletiva, ou seja, relacionada à exposição, a resposta, ou a ambos.

Viés de Informação

- O viés de informação está relacionado à classificação da condição de exposição (expostos / não-expostos) e/ou na determinação da resposta.
- Medidas são aproximações de atributos do mundo real, baseadas em modelos conceituais (teorias).
- Sensibilidade, especificidade, validade, confiabilidade são exemplos de medidas utilizadas para quantificar os erros de classificação.

- Validade Interna: sujeito a confundimento e viés;
- Validade Externa: representatividade da amostra.

Sujeito a critérios de inclusão e exclusão do estudo clínico.

Etapas de uma Análise Estatística

- 1 Descrição/Entendimento da Pergunta Clínica.
- 2 Descrição da Casuística: desenho do estudo e a coleta dos dados.
- 3 Análise Descritiva.
- 4 Análise Univariada.
- 5 Análise Multivariada (na maioria das situações).

Análise Descritiva: Sempre Explorar antes de Analisar.

- 1 Ferramentas: gráficos e tabelas.
- 2 Objetivo:
 - Conhecimento das variáveis envolvidas no estudo.
 - Detectar erros na coleta e digitação dos dados.

Análise Univariada

- 1 Ferramentas: testes t, qui-quadrado, correlação, etc.
- 2 Objetivo:
 - Identificação preliminar de uma possível relação entre o desfecho e cada uma das covariáveis.
 - Identificação de possíveis fatores de confusão.
- 3 Identificar uma possível forma funcional de covariáveis quantitativas;
- 4 Usualmente, não é a análise estatística final.

- 1 Análise confirmatória ou inferencial.
- 2 Ferramentas: modelos estatísticos: linear, logístico, de Cox, etc.
- 3 Objetivo:
 - Análise Estatística conjunta: controla e compara por todas as variáveis medidas no estudo.
 - A conclusão final do estudo é baseado nesta análise.

Como identificar o modelo adequado?

- 1 A resposta (desfecho) define o modelo multivariado para a análise dos dados.
- 2 Modelos para Resposta Quantitativa.
 - Regressão linear múltipla;
 - Modelo de Cox: a resposta é o tempo até a ocorrência de um evento de interesse (presença de censuras).
- 3 Modelos para Resposta Qualitativa ou Contagem.
 - Modelo de Regressão Logística (binária ou politômica);
 - Modelo de Regressão de Poisson: contagem.
 - Outros: modelos beta, gama, etc.

$$\text{Desfecho } (Y) = f(X) + \epsilon(\text{erro})$$

f : modelo verdadeiro (desconhecido) e X as covariáveis medidas.

Objetivo Estatístico: Estimar f obtendo \hat{f} .

Para tal necessitamos de:

- Modelo Estatístico (na maioria das vezes);
- Dados;
- Método estatístico/inferencial.

Temos dois tipos de erros:

- Reduzível: $f - \hat{f}$;
- Irreduzível: ϵ

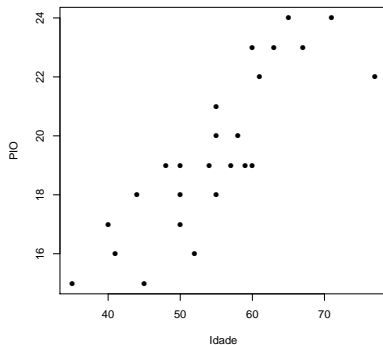
Métodos para estimar f :

- Paramétrico: interpretável e eficiente;
- Não-paramétrico: preciso, flexível e limitado.

* Introduction to Statistical Learning (James et. al, 2013)

Exemplo: Pressão Intra-Ocular (PIO) e idade

- Amostra de 25 pacientes.



- Solução: Encontrar/estimar f

$$\text{PIO} = f(\text{idade}) + \epsilon$$

Exemplo: Pressão Intra-Ocular (PIO) e idade

Possíveis Modelos f

1 Paramétricos

- Modelo mais simples: reta.

$$f(\text{idade}) = \beta_0 + \beta_1 \text{ idade}$$

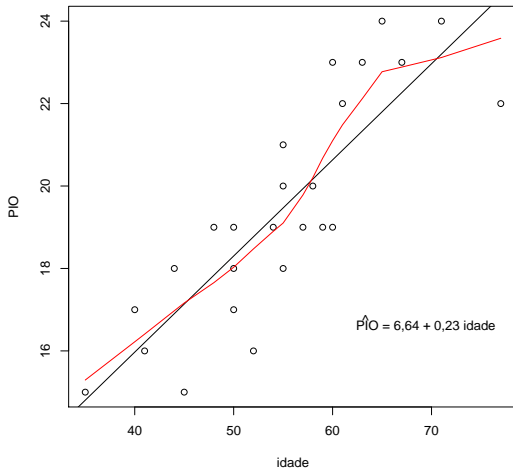
- Modelo quadrático

$$f(\text{idade}) = \beta_0 + \beta_1 \text{ idade} + \beta_2 \text{ idade}^2$$

2 Não-Paramétrico

Por exemplo, LOWESS (suavizador local)

Exemplo: Pressão Intra-Ocular (PIO) e idade



(a) reta de regressão (preta); (b) lowess (vermelha).

Como encontrar f ?

Qual modelo f devemos utilizar?

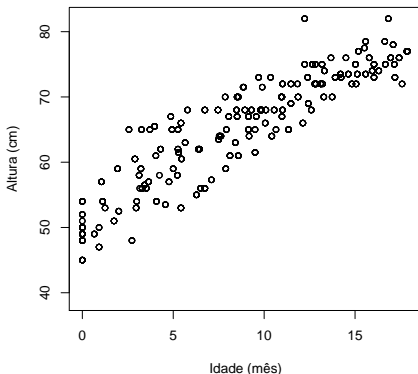
- 1 O objetivo do estudo deve nortear a decisão.
 - Explicar o desfecho (inferência): modelos paramétricos;
 - Experiência anterior;
 - Predição: modelos não-paramétricos.
- 2 Sempre lembrando que:
 - Modelos Paramétricos: interpretável, pouco flexível e eficiente;
 - Métodos Não-paramétricos: preciso, flexível, pouco eficiente e limitado.
 - Modelo semi-paramétrico: híbrido.

Exemplo: Crescimento de crianças infectadas pelo HIV

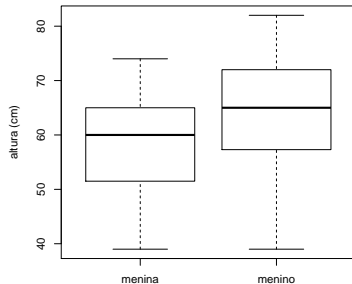
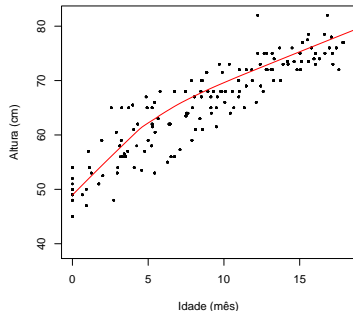
Y altura de crianças entre 0 e 18 meses de idade.

X idade da criança

Casuística Longitudinal: Crianças infectadas pelo HIV (transmissão vertical) do ambulatório de Infectologia Pediátrica da UFMG.



Exemplo: Crescimento de crianças infectadas pelo HIV



Observações:

- Altura média não parece ser linear na idade.
- Sexo da criança explica parte da dispersão dos dados.

1 Caminho convencional:

- Método/Modelo Paramétrico.
- Método Inferencial: Mínimos quadrados/Máxima Verossimilhança
- Ignora casos com dados perdidos.
- Dificuldade em tratar violação de suposições.

2 Métodos Modernos (terceira disciplina de estatística??)

- Combinar Métodos Paramétricos e Não-Paramétricos (splines).
- Tratamento de Dados Perdidos: Imputação.
- Avaliar forma funcional de modelo e de covariáveis contínuas.

3 Necessidade: software estatístico: R