Análise Exploratória e Modelagem Estatística de Dados

Aparecida D. P. Souza
Faculdade de Ciências e Tecnologia - Unesp
www.fct.unesp.br/adps
adps@fct.unesp.br

XIII SEMANA DA ESTATÍSTICA - UEM
4 a 7 de Março de 2017
Maringá/PR

Outlines
Parte I
Parte II
Parte III

Parte I

Outlines
Parte I
Parte II
Parte III

1. Introdução
   - Objetivos
2. Dados Cardiologia
   - Tabelas de Contingência
   - Razão de Chances e Risco Relativo
3. Modelagem Estatística
   - Modelos Lineares Generalizados
   - Regressão Logística
4. Resultados da Modelagem
   - Estimativas para os Parâmetros e Distribuições
   - Estimativas para Razões de Chances e Interpretações
5. Extensões
   - Modelo de Regressão Logística com Efeito Aleatório

Parte II

Outlines
Parte I
Parte II
Parte III

6 Análise do Ganho de Peso em Gestantes

- Motivando Regressão Hierárquica
- Ganho de Peso em Gestantes
  - Modelo hierárquico de dois estágios
  - Modelo hierárquico de três estágios
  - Modelo de Crescimento Linear

Parte III

Outlines
Parte I
Parte II
Parte III

7. Modelos para Dados de Contagem
   - Modelo de Poisson
   - Aplicações
     - Relação entre o hábito de fumar e mortes por doenças coronarianas

Introdução

Introdução
Objetivos
Dados Cardiologia
Tabelas de Contingência
Razão de Chances e Risco Relativo
Modelagem Estatística
Modelos Lineares Generalizados
Regressão Logística
Resultados da Modelagem
Estimativas para os Parâmetros e Distribuições
Estimativas para Razões de Chances e Interpretações
Extensões

Objetivos

- Apresentar e discutir:
  - análise exploratória de alguns conjuntos de dados reais;
  - análise mais elaborada envolvendo modelagem estatística.

- Alguns conjuntos de dados:
  - Análise do risco de óbito hospitalar em pacientes acometidos por infarto agudo do miocárdio;
  - Análise do comportamento do ganho de peso em gestantes ao longo da gestação.

Introdução
Objetivos
Dados Cardiologia
Tabelas de Contingência
Razão de Chances e Risco Relativo
Modelagem Estatística
Modelos Lineares Generalizados
Regressão Logística
Resultados da Modelagem
Estimativas para os Parâmetros e Distribuições
Estimativas para Razões de Chances e Interpretações
Extensões

Objetivos

- Apresentar e discutir:
  - análise exploratória de alguns conjuntos de dados reais;
  - análise mais elaborada envolvendo modelagem estatística.

- Alguns conjuntos de dados:
  - Análise do risco de óbito hospitalar em pacientes acometidos por infarto agudo do miocárdio;
  - Análise do comportamento do ganho de peso em gestantes ao longo da gestação.

Introdução
Objetivos
Dados Cardiologia
Tabelas de Contingência
Razão de Chances e Risco Relativo
Modelagem Estatística
Modelos Lineares Generalizados
Regressão Logística
Resultados da Modelagem
Estimativas para os Parâmetros e Distribuições
Estimativas para Razões de Chances e Interpretações
Extensões

Objetivos

- Alguns conjuntos de dados:
  - Análise da distribuição espaço temporal da mortalidade pelos cânceres de maior importância quantitativa, segundo as microrregiões do Estado de São Paulo;
  - Análise da incidência de malária na Amazonia considerando espaço e tempo;
  - Análise da incidência de alerta de poluição do ar na cidade do México.

Introdução
Objetivos
Dados Cardiologia
Tabelas de Contingência
Razão de Chances e Risco Relativo
Modelagem Estatística
Modelos Lineares Generalizados
Regressão Logística
Resultados da Modelagem
Estimativas para os Parâmetros e Distribuições
Estimativas para Razões de Chances e Interpretações
Extensões

Objetivos

- Alguns conjuntos de dados:
  - Análise da distribuição espaço temporal da mortalidade pelos cânceres de maior importância quantitativa, segundo as microrregiões do Estado de São Paulo;
  - Análise da incidência de malária na Amazonia considerando espaço e tempo;
  - Análise da incidência de alerta de poluição do ar na cidade do México.

Introdução
Objetivos
Dados Cardiologia
Tabelas de Contingência
Razão de Chances e Risco Relativo
Modelagem Estatística
Modelos Lineares Generalizados
Regressão Logística
Resultados da Modelagem
Estimativas para os Parâmetros e Distribuições
Estimativas para Razões de Chances e Interpretações
Extensões

Objetivos

- Alguns conjuntos de dados:
  - Análise da distribuição espaço temporal da mortalidade pelos cânceres de maior importância quantitativa, segundo as microrregiões do Estado de São Paulo;
  - Análise da incidência de malária na Amazonia considerando espaço e tempo;
  - Análise da incidência de alerta de poluição do ar na cidade do México.

Introdução
Objetivos
Dados Cardiologia
Tabelas de Contingência
Razão de Chances e Risco Relativo
Modelagem Estatística
Modelos Lineares Generalizados
Regressão Logística
Resultados da Modelagem
Estimativas para os Parâmetros e Distribuições
Estimativas para Razões de Chances e Interpretações
Extensões

Dados Cardiologia

- Exames iniciais em pacientes com IAM são insuficientes para a predição do risco de óbito, gerando outros riscos e custos hospitalares.
- Características clínicas registradas na admissão

$\Downarrow$

Predição do Risco de óbito hospitalar pós IAM

$\Downarrow$

Tomada de Decisão entre Terapias Alternativas e uso de Recursos Clínicos.

Problema

Predizer óbito hospitalar a partir de variáveis clínicas registradas na admissão do paciente no hospital

Introdução
Objetivos
Dados Cardiologia
Tabelas de Contingência
Razão de Chances e Risco Relativo
Modelagem Estatística
Modelos Lineares Generalizados
Regressão Logística
Resultados da Modelagem
Estimativas para os Parâmetros e Distribuições
Estimativas para Razões de Chances e Interpretações
Extensões

Dados Cardiologia

- Exames iniciais em pacientes com IAM são insuficientes para a predição do risco de óbito, gerando outros riscos e custos hospitalares.
- Características clínicas registradas na admissão

$\Downarrow$

Predição do Risco de óbito hospitalar pós IAM

$\Downarrow$

Tomada de Decisão entre Terapias Alternativas e uso de Recursos Clínicos.

Problema

Predizer óbito hospitalar a partir de variáveis clínicas registradas na admissão do paciente no hospital.

Introdução
Objetivos
Dados Cardiologia
Tabelas de Contingência
Razão de Chances e Risco Relativo
Modelagem Estatística
Modelos Lineares Generalizados
Regressão Logística
Resultados da Modelagem
Estimativas para os Parâmetros e Distribuições
Estimativas para Razões de Chances e Interpretações

Dados Cardiologia

- Exames iniciais em pacientes com IAM são insuficientes para a predição do risco de óbito, gerando outros riscos e custos hospitalares.
- Características clínicas registradas na admissão

$$\Downarrow$$

Predição do Risco de óbito hospitalar pós IAM

$$\Downarrow$$

Tomada de Decisão entre Terapias Alternativas e uso de Recursos Clínicos.

Problema

Predizer óbito hospitalar a partir de variáveis clínicas registradas na admissão do paciente no hospital.

Introdução
Objetivos
Dados Cardiologia
Tabelas de Contingência
Razão de Chances e Risco Relativo
Modelagem Estatística
Modelos Lineares Generalizados
Regressão Logística
Resultados da Modelagem
Estimativas para os Parâmetros e Distribuições
Estimativas para Razões de Chances e Interpretações
Extensões

Dados Cardiologia

- Exames iniciais em pacientes com IAM são insuficientes para a predição do risco de óbito, gerando outros riscos e custos hospitalares.
- Características clínicas registradas na admissão

$$\Downarrow$$

Predição do Risco de óbito hospitalar pós IAM

$$\Downarrow$$

Tomada de Decisão entre Terapias Alternativas e uso de Recursos Clínicos.

Problema

Predizer óbito hospitalar a partir de variáveis clínicas registradas na admissão do paciente no hospital.

Introdução
Objetivos
Dados Cardiologia
Tabelas de Contingência
Razão de Chances e Risco Relativo
Modelagem Estatística
Modelos Lineares Generalizados
Regressão Logística
Resultados da Modelagem
Estimativas para os Parâmetros e Distribuições
Estimativas para Razões de Chances e Interpretações

Dados Cardiologia

- Exames iniciais em pacientes com IAM são insuficientes para a predição do risco de óbito, gerando outros riscos e custos hospitalares.
- Características clínicas registradas na admissão

$$\Downarrow$$

Predição do Risco de óbito hospitalar pós IAM

$$\Downarrow$$

Tomada de Decisão entre Terapias Alternativas e uso de Recursos Clínicos.

Problema

Predizer óbito hospitalar a partir de variáveis clínicas registradas na admissão do paciente no hospital.

Introdução
Objetivos
Dados Cardiologia
Tabelas de Contingência
Razão de Chances e Risco Relativo
Modelagem Estatística
Modelos Lineares Generalizados
Regressão Logística
Resultados da Modelagem
Estimativas para os Parâmetros e Distribuições
Estimativas para Razões de Chances e Interpretações
Extensões

Dados Cardiologia

- Estudo Prospectivo
- Referem-se a 600 pacientes
  Hospital Pró-Cardíaco do Rio de Janeiro
  Janeiro de 91 a Junho de 96
- Principais Variáveis:
  - Idade;
  - Sexo;
  - Hipertensão arterial;
  - Infarto Prévio;
  - Hábito de fumar;
  - Presença de diabetes;
  - Localização do Infarto;
  - Uso de Trombolítico;
  - Intensidade do infarto na admissão (killip de entrada);

Introdução
Objetivos
Dados Cardiologia
Tabelas de Contingência
Razão de Chances e Risco Relativo
Modelagem Estatística
Modelos Lineares Generalizados
Regressão Logística
Resultados da Modelagem
Estimativas para os Parâmetros e Distribuições
Estimativas para Razões de Chances e Interpretações
Extensões

Dados Cardiologia

- Estudo Prospectivo
- Referem-se a 600 pacientes
  Hospital Pró-Cardíaco do Rio de Janeiro
  Janeiro de 91 a Junho de 96
- Principais Variáveis:
  - Idade;
  - Sexo;
  - Hipertensão arterial;
  - Infarto Prévio;
  - Hábito de fumar;
  - Presença de diabetes;
  - Localização do Infarto;
  - Uso de Trombolítico;
  - Intensidade do infarto na admissão (killip de entrada);

Introdução
Objetivos
Dados Cardiologia
Tabelas de Contingência
Razão de Chances e Risco Relativo
Modelagem Estatística
Modelos Lineares Generalizados
Regressão Logística
Resultados da Modelagem
Estimativas para os Parâmetros e Distribuições
Estimativas para Razões de Chances e Interpretações
Extensões

Dados Cardiologia

- Estudo Prospectivo
- Referem-se a 600 pacientes
  Hospital Pró-Cardíaco do Rio de Janeiro
  Janeiro de 91 a Junho de 96
- Principais Variáveis:
  - Idade;
  - Sexo;
  - Hipertensão arterial;
  - Infarto Prévio;
  - Hábito de fumar;
  - Presença de diabetes;
  - Localização do Infarto;
  - Uso de Trombolítico;
  - Intensidade do infarto na admissão (killip de entrada);

Introdução
Objetivos
Dados Cardiologia
Tabelas de Contingência
Razão de Chances e Risco Relativo
Modelagem Estatística
Modelos Lineares Generalizados
Regressão Logística
Resultados da Modelagem
Estimativas para os Parâmetros e Distribuições
Estimativas para Razões de Chances e Interpretações
Extensões

Dados Cardiologia

Caracterização das Variáveis Clínicas em estudo.

Variável	identificação	codificação
óbito hospitalar	**obith**	não = 0, sim = 1
idade	**idade**	anos
sexo	**sexo**	feminino = 0, masculino = 1
hipertensão arterial	**has**	ausência = 0, presença = 1
infarto agudo do miocárdio prévio	**iamp**	não = 0, sim = 1
diabete	**diab**	ausência = 0, presença = 1
fumo	**fumo**	não fumante = 0, fumante = 1
localização: parede anterior	**parea**	não = 0, sim = 1
localização: parede inferior	**parei**	não = 0, sim = 1
localização: parede lateral	**parel**	não = 0, sim = 1
localização: sem onda Q	**pares**	presença = 0; ausência de onda = 1
Killip de entrada (intensidade)	**kient**	1,2,3,4: caso menos grave = 1; caso mais grave = 4

Introdução
Objetivos
Dados Cardiologia
Tabelas de Contingência
Razão de Chances e Risco Relativo
Modelagem Estatística
Modelos Lineares Generalizados
Regressão Logística
Resultados da Modelagem
Estimativas para os Parâmetros e Distribuições
Estimativas para Razões de Chances e Interpretações
Extensões

Dados Cardiologia

Tabela: Distribuição dos Pacientes em função da Idade e variável resposta "Óbito Hospitalar".

Idade	N^o óbitos	N^o pacientes	% óbitos
30 ⊣ 40	1	21	4,8
40 ⊣ 50	3	88	3,4
50 ⊣ 60	6	109	5,5
60 ⊣ 70	18	168	10,7
70 ⊣ 80	31	140	22,1
80 ⊣ 90	19	68	27,9
90 ⊣ 100	3	6	50,0
Total	81	600	13,5

Introdução
Objetivos
Dados Cardiologia
Tabelas de Contingência
Razão de Chances e Risco Relativo
Modelagem Estatística
Modelos Lineares Generalizados
Regressão Logística
Resultados da Modelagem
Estimativas para os Parâmetros e Distribuições
Estimativas para Razões de Chances e Interpretações

Dados Cardiologia

Tabela: Distribuição dos Pacientes em função da Idade e variável resposta "Óbito Hospitalar".

Idade	N^o_ óbitos	N^o_ pacientes	% óbitos
≤ 55	5	160	3,1
$55 \dashv 65$	16	151	10,6
$65 \dashv 75$	24	149	16,1
> 75	36	140	25,7

Tabela: Distribuição dos Pacientes em função do Sexo e da resposta "Óbito Hospitalar".

Sexo	N^o_ óbitos	N^o_ pacientes	% óbitos
Feminino	38	165	23,0
Masculino	43	435	9,9

Introdução
Objetivos
Dados Cardiologia
Tabelas de Contingência
Razão de Chances e Risco Relativo
Modelagem Estatística
Modelos Lineares Generalizados
Regressão Logística
Resultados da Modelagem
Estimativas para os Parâmetros e Distribuições
Estimativas para Razões de Chances e Interpretações

Dados Cardiologia

Tabela: Distribuição dos Pacientes em função da Idade e variável resposta "Óbito Hospitalar".

Idade	N^o óbitos	N^o pacientes	% óbitos
≤ 55	5	160	3,1
$55 \dashv 65$	16	151	10,6
$65 \dashv 75$	24	149	16,1
> 75	36	140	25,7

Tabela: Distribuição dos Pacientes em função do Sexo e da resposta "Óbito Hospitalar".

Sexo	N^o óbitos	N^o pacientes	% óbitos
Feminino	38	165	23,0
Masculino	43	435	9,9

Introdução
Objetivos
Dados Cardiologia
Tabelas de Contingência
Razão de Chances e Risco Relativo
Modelagem Estatística
Modelos Lineares Generalizados
Regressão Logística
Resultados da Modelagem
Estimativas para os Parâmetros e Distribuições
Estimativas para Razões de Chances e Interpretações

Dados Cardiologia

Tabela: Distribuição dos Pacientes em função de Presença/ Ausência de Hipertensão Arterial e da resposta "Óbito Hospitalar".

Hipertensão	N^o_- óbitos	N^o_- pacientes	% óbitos
Ausência	41	311	13,2
Presença	40	289	13,8

Tabela: Distribuição dos Pacientes em função de Infarto Prévio e da resposta "Óbito Hospitalar".

Infarto Prévio	N^o_- óbitos	N^o_- pacientes	% óbitos
Não	52	489	10,6
Sim	29	111	26,1

Introdução
Objetivos
Dados Cardiologia
Tabelas de Contingência
Razão de Chances e Risco Relativo
Modelagem Estatística
Modelos Lineares Generalizados
Regressão Logística
Resultados da Modelagem
Estimativas para os Parâmetros e Distribuições
Estimativas para Razões de Chances e Interpretações
Extensões

Dados Cardiologia

Tabela: Distribuição dos Pacientes em função de Presença/ Ausência de Hipertensão Arterial e da resposta "Óbito Hospitalar".

Hipertensão	N^o_- óbitos	N^o_- pacientes	% óbitos
Ausência	41	311	13,2
Presença	40	289	13,8

Tabela: Distribuição dos Pacientes em função de Infarto Prévio e da resposta "Óbito Hospitalar".

Infarto Prévio	N^o_- óbitos	N^o_- pacientes	% óbitos
Não	52	489	10,6
Sim	29	111	26,1

Introdução
Objetivos
Dados Cardiologia
Tabelas de Contingência
Razão de Chances e Risco Relativo
Modelagem Estatística
Modelos Lineares Generalizados
Regressão Logística
Resultados da Modelagem
Estimativas para os Parâmetros e Distribuições
Estimativas para Razões de Chances e Interpretações
Extensões

Dados Cardiologia

Tabela: Distribuição dos Pacientes em função de Presença/ Ausência de Diabetes e da resposta "Óbito Hospitalar".

Diabetes	N^o óbitos	N^o pacientes	% óbitos
Ausência	63	503	12,5
Presença	18	97	18,6

Tabela: Distribuição dos Pacientes em função do Hábito de Fumar e da resposta "Óbito Hospitalar".

Hábito de Fumar	N^o óbitos	N^o pacientes	% óbitos
Não	66	370	17,8
Sim	15	229	6,5

Introdução
Objetivos
Dados Cardiologia
Tabelas de Contingência
Razão de Chances e Risco Relativo
Modelagem Estatística
Modelos Lineares Generalizados
Regressão Logística
Resultados da Modelagem
Estimativas para os Parâmetros e Distribuições
Estimativas para Razões de Chances e Interpretações

Dados Cardiologia

Tabela: Distribuição dos Pacientes em função de Presença/ Ausência de Diabetes e da resposta "Óbito Hospitalar".

Diabetes	N^o_- óbitos	N^o_- pacientes	% óbitos
Ausência	63	503	12,5
Presença	18	97	18,6

Tabela: Distribuição dos Pacientes em função do Hábito de Fumar e da resposta "Óbito Hospitalar".

Hábito de Fumar	N^o_- óbitos	N^o_- pacientes	% óbitos
Não	66	370	17,8
Sim	15	229	6,5

Introdução
Objetivos
Dados Cardiologia
Tabelas de Contingência
Razão de Chances e Risco Relativo
Modelagem Estatística
Modelos Lineares Generalizados
Regressão Logística
Resultados da Modelagem
Estimativas para os Parâmetros e Distribuições
Estimativas para Razões de Chances e Interpretações
Extensões

Dados Cardiologia

Tabela: Distribuição dos Pacientes em função do Uso de Trombolítico e da resposta "Óbito Hospitalar".

Trombolítico	N^o_- óbitos	N^o_- pacientes	% óbitos
Não	57	356	16,0
Sim	24	239	10,0

Tabela: Distribuição dos Pacientes em função da Intensidade do Infarto (Killip de entrada) e da resposta "Óbito Hospitalar".

Killip	N^o_- óbitos	N^o_- pacientes	% óbitos
1	32	506	6,3
2	25	60	41,0
3	11	20	55,0
4	13	14	92,9

Introdução
Objetivos
Dados Cardiologia
Tabelas de Contingência
Razão de Chances e Risco Relativo
Modelagem Estatística
Modelos Lineares Generalizados
Regressão Logística
Resultados da Modelagem
Estimativas para os Parâmetros e Distribuições
Estimativas para Razões de Chances e Interpretações
Extensões

Dados Cardiologia

Tabela: Distribuição dos Pacientes em função da Localização do IAM e da resposta "Óbito Hospitalar".

Parea	N^o_ óbitos	N^o_ pacientes	% óbitos
Não	40	377	10,6
Sim	41	223	18,4

Tabela: Distribuição dos Pacientes em função da Localização do IAM e da resposta "Óbito Hospitalar".

Parei	N^o_ óbitos	N^o_ pacientes	% óbitos
Não	62	367	16,9
Sim	19	233	8,2

Introdução
Objetivos
Dados Cardiologia
Tabelas de Contingência
Razão de Chances e Risco Relativo
Modelagem Estatística
Modelos Lineares Generalizados
Regressão Logística
Resultados da Modelagem
Estimativas para os Parâmetros e Distribuições
Estimativas para Razões de Chances e Interpretações
Extensões

Dados Cardiologia

Tabela: Distribuição dos Pacientes em função da Localização do IAM e da resposta "Óbito Hospitalar".

Parea	N^o óbitos	N^o pacientes	% óbitos
Não	40	377	10,6
Sim	41	223	18,4

Tabela: Distribuição dos Pacientes em função da Localização do IAM e da resposta "Óbito Hospitalar".

Parei	N^o óbitos	N^o pacientes	% óbitos
Não	62	367	16,9
Sim	19	233	8,2

Introdução
Objetivos
Dados Cardiologia
Tabelas de Contingência
Razão de Chances e Risco Relativo
Modelagem Estatística
Modelos Lineares Generalizados
Regressão Logística
Resultados da Modelagem
Estimativas para os Parâmetros e Distribuições
Estimativas para Razões de Chances e Interpretações
Extensões

Dados Cardiologia

Tabela: Distribuição dos Pacientes em função da Localização do IAM e da resposta "Óbito Hospitalar".

Parel	N^o_ óbitos	N^o_ pacientes	% óbitos
Não	79	583	13,6
Sim	2	17	11,8

Tabela: Distribuição dos Pacientes em função da Localização do IAM e da resposta "Óbito Hospitalar".

Pares	N^o_ óbitos	N^o_ pacientes	% óbitos
Não	62	473	13,1
Sim	19	127	15,0

Introdução
Objetivos
Dados Cardiologia
Tabelas de Contingência
Razão de Chances e Risco Relativo
Modelagem Estatística
Modelos Lineares Generalizados
Regressão Logística
Resultados da Modelagem
Estimativas para os Parâmetros e Distribuições
Estimativas para Razões de Chances e Interpretações

Dados Cardiologia

Tabela: Distribuição dos Pacientes em função da Localização do IAM e da resposta "Óbito Hospitalar".

Parel	N^o_- óbitos	N^o_- pacientes	% óbitos
Não	79	583	13,6
Sim	2	17	11,8

Tabela: Distribuição dos Pacientes em função da Localização do IAM e da resposta "Óbito Hospitalar".

Pares	N^o_- óbitos	N^o_- pacientes	% óbitos
Não	62	473	13,1
Sim	19	127	15,0

Introdução
Objetivos
Dados Cardiologia
Tabelas de Contingência
Razão de Chances e Risco Relativo
Modelagem Estatística
Modelos Lineares Generalizados
Regressão Logística
Resultados da Modelagem
Estimativas para os Parâmetros e Distribuições
Estimativas para Razões de Chances e Interpretações

Tabelas de Contingência

- Procedimentos de Pesquisa dão origem às frequências exibidas em Tabelas 2×2:
  - Ensaios Clínicos;
  - Estudos Prospectivos;
  - Estudos Retrospectivos;
  - Estudos Transversais.
- Procedimentos Estatísticos usuais:
  - Testes de hipóteses (qui-quadrado, exato de Fisher, razão de verossimilhanças, homogeneidade (variáveis independentes), ou distribuições homogêneas entre grupos);
  - Cálculo de Medidas de Associação (se existir, qual o grau de associação?):
  - Razão de Chances (odds ratio)

Introdução
Objetivos
Dados Cardiologia
Tabelas de Contingência
Razão de Chances e Risco Relativo
Modelagem Estatística
Modelos Lineares Generalizados
Regressão Logística
Resultados da Modelagem
Estimativas para os Parâmetros e Distribuições
Estimativas para Razões de Chances e Interpretações
Extensões

Tabelas de Contingência

- Procedimentos de Pesquisa dão origem às frequências exibidas em Tabelas 2×2:
  - Ensaios Clínicos;
  - Estudos Prospectivos;
  - Estudos Retrospectivos;
  - Estudos Transversais.
- Procedimentos Estatísticos usuais:
  - Testes de hipóteses (qui-quadrado de Pearson, homogeneidade (para duas ou mais populações independentes), distribuições homogêneas entre grupos);
  - Cálculo de Medidas de Associação (se existe, qual o grau de associação);
  - Razão de Chances (odds ratio)

Introdução
Objetivos
Dados Cardiologia
Tabelas de Contingência
Razão de Chances e Risco Relativo
Modelagem Estatística
Modelos Lineares Generalizados
Regressão Logística
Resultados da Modelagem
Estimativas para os Parâmetros e Distribuições
Estimativas para Razões de Chances e Interpretações
Extensões

Tabelas de Contingência

- Procedimentos de Pesquisa dão origem às frequências exibidas em Tabelas 2×2:
  - Ensaios Clínicos;
  - Estudos Prospectivos;
  - Estudos Retrospectivos;
  - Estudos Transversais.
- Procedimentos Estatísticos usuais:
  - Testes de Hipóteses (qui-quadrado, razão de verossimilhanças, homogeneidade (para várias populações), distribuições homogêneas entre grupos);
  - Cálculo de Medidas de Associação (se existe, qual o grau de associação):
    - Razão de Chances (odds ratio)

Introdução
Objetivos
Dados Cardiologia
Tabelas de Contingência
Razão de Chances e Risco Relativo
Modelagem Estatística
Modelos Lineares Generalizados
Regressão Logística
Resultados da Modelagem
Estimativas para os Parâmetros e Distribuições
Estimativas para Razões de Chances e Interpretações
Extensões

Tabelas de Contingência

- Procedimentos de Pesquisa dão origem às frequências exibidas em Tabelas 2×2:
  - Ensaios Clínicos;
  - Estudos Prospectivos;
  - Estudos Retrospectivos;
  - Estudos Transversais.
- Procedimentos Estatísticos usuais:

Introdução
Objetivos
Dados Cardiologia
Tabelas de Contingência
Razão de Chances e Risco Relativo
Modelagem Estatística
Modelos Lineares Generalizados
Regressão Logística
Resultados da Modelagem
Estimativas para os Parâmetros e Distribuições
Estimativas para Razões de Chances e Interpretações

Tabelas de Contingência

- Procedimentos de Pesquisa dão origem às frequências exibidas em Tabelas 2×2:
  - Ensaios Clínicos;
  - Estudos Prospectivos;
  - Estudos Retrospectivos;
  - Estudos Transversais.
- Procedimentos Estatísticos usuais:
  - O teste Z: utilizado para testar igualdade de probabilidades em dois grupos;
  - O teste χ^2: utilizado para testar independência ou homogeneidade (variáveis independentes ou homogêneas?);
  - O teste exato de Fisher: utilizado para testar independência ou homogeneidade (variáveis independentes ou homogêneas?) quando as suposições para o teste χ^2 não são satisfeitas;
  - Razão de Chances (Odds Ratio).

Introdução
Objetivos
Dados Cardiologia
Tabelas de Contingência
Razão de Chances e Risco Relativo
Modelagem Estatística
Modelos Lineares Generalizados
Regressão Logística
Resultados da Modelagem
Estimativas para os Parâmetros e Distribuições
Estimativas para Razões de Chances e Interpretações

Tabelas de Contingência

- Procedimentos de Pesquisa dão origem às frequências exibidas em Tabelas 2×2:
  - Ensaios Clínicos;
  - Estudos Prospectivos;
  - Estudos Retrospectivos;
  - Estudos Transversais.
- Procedimentos Estatísticos usuais:
  - O teste Z: utilizado para testar igualdade de probabilidades em dois grupos;
  - O teste χ^2: utilizado para testar independência ou homogeneidade (variáveis independentes ou distribuição homogênea entre grupos);
  - Cálculo de Medidas de Associação: se existe, qual o grau da associação;
  - Razão de Chances (odds ratio).

Introdução
Objetivos
Dados Cardiologia
Tabelas de Contingência
Razão de Chances e Risco Relativo
Modelagem Estatística
Modelos Lineares Generalizados
Regressão Logística
Resultados da Modelagem
Estimativas para os Parâmetros e Distribuições
Estimativas para Razões de Chances e Interpretações
Extensões

Tabelas de Contingência

- Procedimentos de Pesquisa dão origem às frequências exibidas em Tabelas 2×2:
  - Ensaios Clínicos;
  - Estudos Prospectivos;
  - Estudos Retrospectivos;
  - Estudos Transversais.
- Procedimentos Estatísticos usuais:
  - O teste Z: utilizado para testar igualdade de probabilidades em dois grupos;
  - O teste χ^2: utilizado para testar independência ou homogeneidade (variáveis independentes ou distribuição homogênea entre grupos);
  - Cálculo de Medidas de Associação: se existe, qual o grau da associação;
  - Razão de Chances (odds ratio)

Introdução
Objetivos
Dados Cardiologia
Tabelas de Contingência
Razão de Chances e Risco Relativo
Modelagem Estatística
Modelos Lineares Generalizados
Regressão Logística
Resultados da Modelagem
Estimativas para os Parâmetros e Distribuições
Estimativas para Razões de Chances e Interpretações
Extensões

Tabelas de Contingência

- Procedimentos de Pesquisa dão origem às frequências exibidas em Tabelas 2×2:
  - Ensaios Clínicos;
  - Estudos Prospectivos;
  - Estudos Retrospectivos;
  - Estudos Transversais.
- Procedimentos Estatísticos usuais:
  - O teste Z: utilizado para testar igualdade de probabilidades em dois grupos;
  - O teste χ^2: utilizado para testar independência ou homogeneidade (variáveis independentes ou distribuição homogênea entre grupos);
  - Cálculo de Medidas de Associação: se existe, qual o grau da associação;
  - Razão de Chances (odds ratio)

Introdução
Objetivos
Dados Cardiologia
Tabelas de Contingência
Razão de Chances e Risco Relativo
Modelagem Estatística
Modelos Lineares Generalizados
Regressão Logística
Resultados da Modelagem
Estimativas para os Parâmetros e Distribuições
Estimativas para Razões de Chances e Interpretações
Extensões

Tabelas de Contingência

- Procedimentos de Pesquisa dão origem às frequências exibidas em Tabelas 2×2:
  - Ensaios Clínicos;
  - Estudos Prospectivos;
  - Estudos Retrospectivos;
  - Estudos Transversais.
- Procedimentos Estatísticos usuais:
  - O teste Z: utilizado para testar igualdade de probabilidades em dois grupos;
  - O teste χ^2: utilizado para testar independência ou homogeneidade (variáveis independentes ou distribuição homogênea entre grupos);
  - Cálculo de Medidas de Associação: se existe, qual o grau da associação;
  - Razão de Chances (odds ratio)

Introdução
Objetivos
Dados Cardiologia
Tabelas de Contingência
Razão de Chances e Risco Relativo
Modelagem Estatística
Modelos Lineares Generalizados
Regressão Logística
Resultados da Modelagem
Estimativas para os Parâmetros e Distribuições
Estimativas para Razões de Chances e Interpretações
Extensões

Tabelas de Contingência

- Procedimentos de Pesquisa dão origem às frequências exibidas em Tabelas 2×2:
  - Ensaios Clínicos;
  - Estudos Prospectivos;
  - Estudos Retrospectivos;
  - Estudos Transversais.
- Procedimentos Estatísticos usuais:
  - O teste Z: utilizado para testar igualdade de probabilidades em dois grupos;
  - O teste χ^2: utilizado para testar independência ou homogeneidade (variáveis independentes ou distribuição homogênea entre grupos);
  - Cálculo de Medidas de Associação: se existe, qual o grau da associação;
  - Razão de Chances (odds ratio)

Introdução
Objetivos
Dados Cardiologia
Tabelas de Contingência
Razão de Chances e Risco Relativo
Modelagem Estatística
Modelos Lineares Generalizados
Regressão Logística
Resultados da Modelagem
Estimativas para os Parâmetros e Distribuições
Estimativas para Razões de Chances e Interpretações
Extensões

Tabelas de Contingência

- Para o cálculo da Razão de Chances (ψ) considere dois grupos de pessoas e as proporções:
  - p_i de casos com determinado resultado no grupo $i = 1, 2$;
  - $q_i = 1 - p_i$ de casos em que o resultado não ocorreu no grupo $i = 1, 2$;

- A chance associada ao resultado em questão é

$\omega_1 = \frac{p_1}{q_1}$ no grupo I, $\omega_2 = \frac{p_2}{q_2}$ no grupo II

$$RC = \frac{\omega_1}{\omega_2}$$

Introdução
Objetivos
Dados Cardiologia
Tabelas de Contingência
Razão de Chances e Risco Relativo
Modelagem Estatística
Modelos Lineares Generalizados
Regressão Logística
Resultados da Modelagem
Estimativas para os Parâmetros e Distribuições
Estimativas para Razões de Chances e Interpretações
Extensões

Tabelas de Contingência

- Para o cálculo da Razão de Chances (ψ) considere dois grupos de pessoas e as proporções:
  - p_i de casos com determinado resultado no grupo $i = 1, 2$;
  - $q_i = 1 - p_i$ de casos em que o resultado não ocorreu no grupo $i = 1, 2$;

- A chance associada ao resultado em questão é

$\omega_1 = \frac{p_1}{q_1}$ no grupo I, $\omega_2 = \frac{p_2}{q_2}$ no grupo II

$$RC = \frac{\omega_1}{\omega_2}$$

Introdução
Objetivos
Dados Cardiologia
Tabelas de Contingência
Razão de Chances e Risco Relativo
Modelagem Estatística
Modelos Lineares Generalizados
Regressão Logística
Resultados da Modelagem
Estimativas para os Parâmetros e Distribuições
Estimativas para Razões de Chances e Interpretações
Extensões

Tabelas de Contingência

Tabela: Distribuição dos Pacientes em função de Infarto Prévio e da resposta "Óbito Hospitalar".

Infarto Prévio	N^o_- óbitos	N^o_- pacientes	% óbitos
Não	52	489	10,6
Sim	29	111	26,1

$\omega_1 = \frac{\frac{29}{111}}{\frac{82}{111}}$ no grupo I, $\omega_2 = \frac{\frac{52}{489}}{\frac{437}{489}}$ no grupo II

$RC = \frac{\omega_1}{\omega_2} = 2,97$

$RR = \frac{0,261}{0,106} = 2,46$ (Risco Relativo).

Introdução
Objetivos
Dados Cardiologia
Tabelas de Contingência
Razão de Chances e Risco Relativo
Modelagem Estatística
Modelos Lineares Generalizados
Regressão Logística
Resultados da Modelagem
Estimativas para os Parâmetros e Distribuições
Estimativas para Razões de Chances e Interpretações

Modelagem Estatística

Introdução
Objetivos
Dados Cardiologia
Tabelas de Contingência
Razão de Chances e Risco Relativo
Modelagem Estatística
Modelos Lineares Generalizados
Regressão Logística
Resultados da Modelagem
Estimativas para os Parâmetros e Distribuições
Estimativas para Razões de Chances e Interpretações
Extensões

Modelos de Regressão

- Um dos mais importantes problemas em Estatística é o de modelar a relação entre respostas de interesse e variáveis associadas.
- Análise de regressão
- Variável resposta contínua
- Variável resposta discreta

Introdução
Objetivos
Dados Cardiologia
Tabelas de Contingência
Razão de Chances e Risco Relativo
Modelagem Estatística
Modelos Lineares Generalizados
Regressão Logística
Resultados da Modelagem
Estimativas para os Parâmetros e Distribuições
Estimativas para Razões de Chances e Interpretações
Extensões

Modelos de Regressão

- Um dos mais importantes problemas em Estatística é o de modelar a relação entre respostas de interesse e variáveis associadas.
- Análise de regressão
- Variável resposta contínua
  - distribuição normal;
  - dispersão em torno do valor médio constante (homocedasticidade);
  - ...
- Variável resposta discreta

Introdução
Objetivos
Dados Cardiologia
Tabelas de Contingência
Razão de Chances e Risco Relativo
Modelagem Estatística
Modelos Lineares Generalizados
Regressão Logística
Resultados da Modelagem
Estimativas para os Parâmetros e Distribuições
Estimativas para Razões de Chances e Interpretações
Extensões

Modelos de Regressão

- Um dos mais importantes problemas em Estatística é o de modelar a relação entre respostas de interesse e variáveis associadas.
- Análise de regressão
- Variável resposta contínua
  - distribuição normal;
  - dispersão em torno do valor médio constante (homocedasticidade);
  - independência entre as observações.
- Variável resposta discreta

Introdução
Objetivos
Dados Cardiologia
Tabelas de Contingência
Razão de Chances e Risco Relativo
Modelagem Estatística
Modelos Lineares Generalizados
Regressão Logística
Resultados da Modelagem
Estimativas para os Parâmetros e Distribuições
Estimativas para Razões de Chances e Interpretações
Extensões

Modelos de Regressão

- Um dos mais importantes problemas em Estatística é o de modelar a relação entre respostas de interesse e variáveis associadas.
- Análise de regressão
- Variável resposta contínua
  - distribuição normal;
  - dispersão em torno do valor médio constante (homocedasticidade);
  - independência entre as observações.
- Variável resposta discreta

Introdução
Objetivos
Dados Cardiologia
Tabelas de Contingência
Razão de Chances e Risco Relativo
Modelagem Estatística
Modelos Lineares Generalizados
Regressão Logística
Resultados da Modelagem
Estimativas para os Parâmetros e Distribuições
Estimativas para Razões de Chances e Interpretações
Extensões

Modelos de Regressão

- Um dos mais importantes problemas em Estatística é o de modelar a relação entre respostas de interesse e variáveis associadas.
- Análise de regressão
- Variável resposta contínua
  - distribuição normal;
  - dispersão em torno do valor médio constante (homocedasticidade);
  - independência entre as observações.
- Variável resposta discreta

Introdução
Objetivos
Dados Cardiologia
Tabelas de Contingência
Razão de Chances e Risco Relativo
Modelagem Estatística
Modelos Lineares Generalizados
Regressão Logística
Resultados da Modelagem
Estimativas para os Parâmetros e Distribuições
Estimativas para Razões de Chances e Interpretações
Extensões

Modelos de Regressão

- Um dos mais importantes problemas em Estatística é o de modelar a relação entre respostas de interesse e variáveis associadas.
- Análise de regressão
- Variável resposta contínua
  - distribuição normal;
  - dispersão em torno do valor médio constante (homocedasticidade);
  - independência entre as observações.
- Variável resposta discreta

Introdução
Objetivos
Dados Cardiologia
Tabelas de Contingência
Razão de Chances e Risco Relativo
Modelagem Estatística
Modelos Lineares Generalizados
Regressão Logística
Resultados da Modelagem
Estimativas para os Parâmetros e Distribuições
Estimativas para Razões de Chances e Interpretações
Extensões

Modelos de Regressão

- Um dos mais importantes problemas em Estatística é o de modelar a relação entre respostas de interesse e variáveis associadas.
- Análise de regressão
- Variável resposta contínua
  - distribuição normal;
  - dispersão em torno do valor médio constante (homocedasticidade);
  - independência entre as observações.
- Variável resposta discreta

Introdução
Objetivos
Dados Cardiologia
Tabelas de Contingência
Razão de Chances e Risco Relativo
Modelagem Estatística
Modelos Lineares Generalizados
Regressão Logística
Resultados da Modelagem
Estimativas para os Parâmetros e Distribuições
Estimativas para Razões de Chances e Interpretações
Extensões

Modelos de Regressão Linear

- Considere o modelo linear normal

$$\begin{aligned} Y_i|\mu_i,\sigma^2 &\sim N(\mu_i,\sigma^2) \\ \mu_i &= \mathbf{x}_i'\beta \end{aligned}$$

  - Y_i variável resposta para a i-ésima observação;
  - μ_i respectiva média, função do preditor linear $\mathbf{x}_i'\beta$;
  - $\mathbf{x}_i' = (1, x_{i1}, \cdots, x_{ik})$ o vetor das covariáveis;
  - β vetor dos parâmetros do modelo;
  - σ^2 a variância observacional, $i = 1, 2, \cdots, n$.

Introdução
Objetivos
Dados Cardiologia
Tabelas de Contingência
Razão de Chances e Risco Relativo
Modelagem Estatística
Modelos Lineares Generalizados
Regressão Logística
Resultados da Modelagem
Estimativas para os Parâmetros e Distribuições
Estimativas para Razões de Chances e Interpretações
Extensões

Modelos de Regressão Linear

- Considere o modelo linear normal

$$\begin{aligned} Y_i|\mu_i,\sigma^2 &\sim N(\mu_i,\sigma^2) \\ \mu_i &= \mathbf{x}_i'\beta \end{aligned}$$

  - Y_i variável resposta para a i-ésima observação;
  - μ_i respectiva média, função do preditor linear $\mathbf{x}_i'\beta$;
  - $\mathbf{x}_i' = (1, x_{i1}, \cdots, x_{ik})$ o vetor das covariáveis;
  - β vetor dos parâmetros do modelo;
  - σ^2 a variância observacional, $i = 1, 2, \cdots, n$.

Introdução
Objetivos
Dados Cardiologia
Tabelas de Contingência
Razão de Chances e Risco Relativo
Modelagem Estatística
Modelos Lineares Generalizados
Regressão Logística
Resultados da Modelagem
Estimativas para os Parâmetros e Distribuições
Estimativas para Razões de Chances e Interpretações
Extensões

Modelos de Regressão Linear

- Considere o modelo linear normal

$$\begin{aligned} Y_i|\mu_i,\sigma^2 &\sim N(\mu_i,\sigma^2) \\ \mu_i &= \mathbf{x}_i'\beta \end{aligned}$$

  - Y_i variável resposta para a i-ésima observação;
  - μ_i respectiva média, função do preditor linear $\mathbf{x}_i'\beta$;
  - $\mathbf{x}_i' = (1, x_{i1}, \cdots, x_{ik})$ o vetor das covariáveis;
  - β vetor dos parâmetros do modelo;
  - σ^2 a variância observacional, $i = 1, 2, \cdots, n$.

Introdução
Objetivos
Dados Cardiologia
Tabelas de Contingência
Razão de Chances e Risco Relativo
Modelagem Estatística
Modelos Lineares Generalizados
Regressão Logística
Resultados da Modelagem
Estimativas para os Parâmetros e Distribuições
Estimativas para Razões de Chances e Interpretações
Extensões

Modelos de Regressão Linear

- Considere o modelo linear normal

$$\begin{aligned} Y_i|\mu_i,\sigma^2 &\sim N(\mu_i,\sigma^2) \\ \mu_i &= \mathbf{x}_i'\beta \end{aligned}$$

  - Y_i variável resposta para a i-ésima observação;
  - μ_i respectiva média, função do preditor linear $\mathbf{x}_i'\beta$;
  - $\mathbf{x}_i' = (1, x_{i1}, \cdots, x_{ik})$ o vetor das covariáveis;
  - β vetor dos parâmetros do modelo;
  - σ^2 a variância observacional, $i = 1, 2, \cdots, n$.

Introdução
Objetivos
Dados Cardiologia
Tabelas de Contingência
Razão de Chances e Risco Relativo
Modelagem Estatística
Modelos Lineares Generalizados
Regressão Logística
Resultados da Modelagem
Estimativas para os Parâmetros e Distribuições
Estimativas para Razões de Chances e Interpretações
Extensões

Modelos de Regressão Linear

- Considere o modelo linear normal

$$\begin{aligned} Y_i|\mu_i,\sigma^2 &\sim N(\mu_i,\sigma^2) \\ \mu_i &= \mathbf{x}_i'\beta \end{aligned}$$

  - Y_i variável resposta para a i-ésima observação;
  - μ_i respectiva média, função do preditor linear $\mathbf{x}_i'\beta$;
  - $\mathbf{x}_i' = (1, x_{i1}, \cdots, x_{ik})$ o vetor das covariáveis;
  - β vetor dos parâmetros do modelo;
  - σ^2 a variância observacional, $i = 1, 2, \cdots, n$.

Introdução
Objetivos
Dados Cardiologia
Tabelas de Contingência
Razão de Chances e Risco Relativo
Modelagem Estatística
Modelos Lineares Generalizados
Regressão Logística
Resultados da Modelagem
Estimativas para os Parâmetros e Distribuições
Estimativas para Razões de Chances e Interpretações
Extensões

Modelos de Regressão Linear

- Considere o modelo linear normal

$$\begin{aligned} Y_i|\mu_i,\sigma^2 &\sim N(\mu_i,\sigma^2) \\ \mu_i &= \mathbf{x}_i'\beta \end{aligned}$$

  - Y_i variável resposta para a i-ésima observação;
  - μ_i respectiva média, função do preditor linear $\mathbf{x}_i'\beta$;
  - $\mathbf{x}_i' = (1, x_{i1}, \cdots, x_{ik})$ o vetor das covariáveis;
  - β vetor dos parâmetros do modelo;
  - σ^2 a variância observacional, $i = 1, 2, \cdots, n$.

Introdução
Objetivos
Dados Cardiologia
Tabelas de Contingência
Razão de Chances e Risco Relativo
Modelagem Estatística
Modelos Lineares Generalizados
Regressão Logística
Resultados da Modelagem
Estimativas para os Parâmetros e Distribuições
Estimativas para Razões de Chances e Interpretações
Extensões

Modelos de Regressão Linear

$$\begin{aligned} Y_i|\mu_i, \sigma^2 &\sim N(\mu_i, \sigma^2) \\ \mu_i &= \mathbf{x}_i'\beta \end{aligned}$$

- 1^a_- equação: distribuição das observações (família de distribuições normais);
- 2^a_- equação: relação entre as médias das observações e a estrutura de explicação do modelo (função de ligação);
- não existe razão para se restringir a classe de distribuições à família normal e a função de ligação à identidade, isto é $g(\mu_i) = \mu_i = \mathbf{x}_i'\beta$.

NELDER, J. A.; WEDDERBURN, R. W. M. Generalized linear models. Journal of the Royal Statistical Society. Series A (General) , Wiley for the Royal Statistical Society, v. 135, n. 3, p. pp. 370384, 1972.

Introdução
Objetivos
Dados Cardiologia
Tabelas de Contingência
Razão de Chances e Risco Relativo
Modelagem Estatística
Modelos Lineares Generalizados
Regressão Logística
Resultados da Modelagem
Estimativas para os Parâmetros e Distribuições
Estimativas para Razões de Chances e Interpretações
Extensões

Modelos de Regressão Linear

$$\begin{aligned} Y_i|\mu_i, \sigma^2 &\sim N(\mu_i, \sigma^2) \\ \mu_i &= \mathbf{x}_i'\beta \end{aligned}$$

- 1^a_- equação: distribuição das observações (família de distribuições normais);
- 2^a_- equação: relação entre as médias das observações e a estrutura de explicação do modelo (função de ligação);
- não existe razão para se restringir a classe de distribuições à família normal e a função de ligação à identidade, isto é $g(\mu_i) = \mu_i = \mathbf{x}_i'\beta$.

NELDER, J. A.; WEDDERBURN, R. W. M. Generalized linear models. Journal of the Royal Statistical Society. Series A (General) , Wiley for the Royal Statistical Society, v. 135, n. 3, p. pp. 370384, 1972.

Introdução
Objetivos
Dados Cardiologia
Tabelas de Contingência
Razão de Chances e Risco Relativo
Modelagem Estatística
Modelos Lineares Generalizados
Regressão Logística
Resultados da Modelagem
Estimativas para os Parâmetros e Distribuições
Estimativas para Razões de Chances e Interpretações
Extensões

Modelos de Regressão Linear

$$\begin{aligned} Y_i|\mu_i, \sigma^2 &\sim N(\mu_i, \sigma^2) \\ \mu_i &= \mathbf{x}_i'\beta \end{aligned}$$

- 1^a_- equação: distribuição das observações (família de distribuições normais);
- 2^a_- equação: relação entre as médias das observações e a estrutura de explicação do modelo (função de ligação);
- não existe razão para se restringir a classe de distribuições à família normal e a função de ligação à identidade, isto é $g(\mu_i) = \mu_i = \mathbf{x}_i'\beta$.

NELDER, J. A.; WEDDERBURN, R. W. M. Generalized linear models. Journal of the Royal Statistical Society. Series A (General) , Wiley for the Royal Statistical Society, v. 135, n. 3, p. pp. 370384, 1972.

Introdução
Objetivos
Dados Cardiologia
Tabelas de Contingência
Razão de Chances e Risco Relativo
Modelagem Estatística
Modelos Lineares Generalizados
Regressão Logística
Resultados da Modelagem
Estimativas para os Parâmetros e Distribuições
Estimativas para Razões de Chances e Interpretações
Extensões

Modelos Lineares Generalizados

- Modelos lineares generalizados
  - Extensão dos modelos lineares normais;
  - Distribuições probabilísticas na família exponencial;
  - Extensão permitiu uma unificação dos procedimentos de inferência tanto para a análise de dados contínuos como discretos.
- Modelo de regressão logística
- Estimação dos Parâmetros
  - Abordagem clássica: Estimativas de máxima verossimilhança (resultados assintóticos, dados amostrais);
  - Abordagem Bayesiana: combina informações contidas da amostra e informações disponíveis sobre o processo através do teorema de Bayes, interesse

Introdução
Objetivos
Dados Cardiologia
Tabelas de Contingência
Razão de Chances e Risco Relativo
Modelagem Estatística
Modelos Lineares Generalizados
Regressão Logística
Resultados da Modelagem
Estimativas para os Parâmetros e Distribuições
Estimativas para Razões de Chances e Interpretações
Extensões

Modelos Lineares Generalizados

- Modelos lineares generalizados
  - Extensão dos modelos lineares normais;
  - Distribuições probabilísticas na família exponencial;
  - Extensão permitiu uma unificação dos procedimentos de inferência tanto para a análise de dados contínuos como discretos.
- Modelo de regressão logística
- Estimação dos Parâmetros
  - Abordagem clássica: Estimadores de Máxima Verossimilhança (Resultados assintóticos, baseado em dados amostrais);
  - Abordagem Bayesiana: combina informações oriundas da amostra e informações disponíveis sobre o processo através do teorema de Bayes, interesse

Introdução
Objetivos
Dados Cardiologia
Tabelas de Contingência
Razão de Chances e Risco Relativo
Modelagem Estatística
Modelos Lineares Generalizados
Regressão Logística
Resultados da Modelagem
Estimativas para os Parâmetros e Distribuições
Estimativas para Razões de Chances e Interpretações
Extensões

Modelos Lineares Generalizados

- Modelos lineares generalizados
  - Extensão dos modelos lineares normais;
  - Distribuições probabilísticas na família exponencial;
  - Extensão permitiu uma unificação dos procedimentos de inferência tanto para a análise de dados contínuos como discretos.
- Modelo de regressão logística
- Estimação dos Parâmetros
  - Abordagens clássicas (estimadores de máxima verossimilhança, testes de hipóteses e intervalos de confiança baseados apenas nas informações dadas pela amostra);
  - Abordagem Bayesiana, combina as informações contidas na amostra e informações disponíveis sobre o processo através do teorema de Bayes, interesse

Introdução
Objetivos
Dados Cardiologia
Tabelas de Contingência
Razão de Chances e Risco Relativo
Modelagem Estatística
Modelos Lineares Generalizados
Regressão Logística
Resultados da Modelagem
Estimativas para os Parâmetros e Distribuições
Estimativas para Razões de Chances e Interpretações

Modelos Lineares Generalizados

- Modelos lineares generalizados
  - Extensão dos modelos lineares normais;
  - Distribuições probabilísticas na família exponencial;
  - Extensão permitiu uma unificação dos procedimentos de inferência tanto para a análise de dados contínuos como discretos.
- Modelo de regressão logística
- Estimação dos Parâmetros

Introdução
Objetivos
Dados Cardiologia
Tabelas de Contingência
Razão de Chances e Risco Relativo
Modelagem Estatística
Modelos Lineares Generalizados
Regressão Logística
Resultados da Modelagem
Estimativas para os Parâmetros e Distribuições
Estimativas para Razões de Chances e Interpretações

Modelos Lineares Generalizados

- Modelos lineares generalizados
  - Extensão dos modelos lineares normais;
  - Distribuições probabilísticas na família exponencial;
  - Extensão permitiu uma unificação dos procedimentos de inferência tanto para a análise de dados contínuos como discretos.
- Modelo de regressão logística
- Estimação dos Parâmetros
  - Abordagem clássica: Estimador de Máxima Verossimilhança (resultados são baseados somente nos dados amostrais).

Introdução
Objetivos
Dados Cardiologia
Tabelas de Contingência
Razão de Chances e Risco Relativo
Modelagem Estatística
Modelos Lineares Generalizados
Regressão Logística
Resultados da Modelagem
Estimativas para os Parâmetros e Distribuições
Estimativas para Razões de Chances e Interpretações

Modelos Lineares Generalizados

- Modelos lineares generalizados
  - Extensão dos modelos lineares normais;
  - Distribuições probabilísticas na família exponencial;
  - Extensão permitiu uma unificação dos procedimentos de inferência tanto para a análise de dados contínuos como discretos.
- Modelo de regressão logística
- Estimação dos Parâmetros
  - Abordagem clássica: Estimador de Máxima Verossimilhança (resultados são baseados somente nos dados amostrais).
  - Abordagem Bayesiana: combina informações oriundas da amostra e informações disponíveis sobre o processo, através do teorema de Bayes. interesse.

Introdução
Objetivos
Dados Cardiologia
Tabelas de Contingência
Razão de Chances e Risco Relativo
Modelagem Estatística
Modelos Lineares Generalizados
Regressão Logística
Resultados da Modelagem
Estimativas para os Parâmetros e Distribuições
Estimativas para Razões de Chances e Interpretações

Modelos Lineares Generalizados

- Modelos lineares generalizados
  - Extensão dos modelos lineares normais;
  - Distribuições probabilísticas na família exponencial;
  - Extensão permitiu uma unificação dos procedimentos de inferência tanto para a análise de dados contínuos como discretos.
- Modelo de regressão logística
- Estimação dos Parâmetros
  - Abordagem clássica: Estimador de Máxima Verossimilhança (resultados são baseados somente nos dados amostrais).
  - Abordagem Bayesiana: combina informações oriundas da amostra e informações disponíveis sobre o processo, através do teorema de Bayes. interesse.

Introdução
Objetivos
Dados Cardiologia
Tabelas de Contingência
Razão de Chances e Risco Relativo
Modelagem Estatística
Modelos Lineares Generalizados
Regressão Logística
Resultados da Modelagem
Estimativas para os Parâmetros e Distribuições
Estimativas para Razões de Chances e Interpretações
Extensões

Modelos Lineares Generalizados

- Modelos lineares generalizados
  - Extensão dos modelos lineares normais;
  - Distribuições probabilísticas na família exponencial;
  - Extensão permitiu uma unificação dos procedimentos de inferência tanto para a análise de dados contínuos como discretos.
- Modelo de regressão logística
- Estimação dos Parâmetros
  - Abordagem clássica: Estimador de Máxima Verossimilhança (resultados são baseados somente nos dados amostrais).
  - Abordagem Bayesiana: combina informações oriundas da amostra e informações disponíveis sobre o processo, através do teorema de Bayes. interesse.

Introdução
Objetivos
Dados Cardiologia
Tabelas de Contingência
Razão de Chances e Risco Relativo
Modelagem Estatística
Modelos Lineares Generalizados
Regressão Logística
Resultados da Modelagem
Estimativas para os Parâmetros e Distribuições
Estimativas para Razões de Chances e Interpretações

Modelo de Regressão para dados binários:

$$Y_i|\pi_i \sim Ber(\pi_i), \quad \pi_i = P(Y_i = 1) = F(\mathbf{x}_i'\boldsymbol{\beta}), \quad i = 1, ..., n$$

- Para o i-ésimo indivíduo:
  - $Y_i = 1$ se a resposta de interesse é observada e $Y_i = 0$ caso contrário;
  - π_i a probabilidade de apresentar a resposta de interesse, $\pi' = (\pi_1, \pi_2, \cdots, \pi_n)$;
  - $\beta = (\beta_0, \beta_1, \beta_2, \cdots, \beta_k)'$ o vetor de parâmetros;
  - $\mathbf{x}_i = (1, x_{i1}, x_{i2}, ..., x_{ik})'$ o vetor das covariáveis.
- F uma função distribuição acumulada que transforma $\mathbf{x}_i'\beta$ para o intervalo desejado.

Introdução
Objetivos
Dados Cardiologia
Tabelas de Contingência
Razão de Chances e Risco Relativo
Modelagem Estatística
Modelos Lineares Generalizados
Regressão Logística
Resultados da Modelagem
Estimativas para os Parâmetros e Distribuições
Estimativas para Razões de Chances e Interpretações
Extensões

Modelos Lineares Generalizados

- As especificações mais comuns para F incluem:

$$F(\mathbf{x}_i'\beta) = \begin{cases} \exp(\mathbf{x}_i'\beta)/[1+\exp(\mathbf{x}_i'\beta)] & \text{logística} \\ \Phi(\mathbf{x}_i'\beta) & \text{probito} \\ 1-\exp[-\exp(\mathbf{x}_i'\beta)] & \text{log log complementar} \end{cases}$$

- Especificando F obtém-se a função de ligação como

$$g(\pi_i) = F^{-1}(\pi_i) = \mathbf{x}_i'\beta.$$

- Com respectivas funções de ligação:

$$g(\pi_i) = \begin{cases} \log\left(\frac{\pi_i}{1-\pi_i}\right), & \text{logística} \\ \Phi^{-1}(\pi_i), & \text{probito} \\ \log(-\log(1-\pi_i)), & \text{log log complementar.} \end{cases}$$

Introdução
Objetivos
Dados Cardiologia
Tabelas de Contingência
Razão de Chances e Risco Relativo
Modelagem Estatística
Modelos Lineares Generalizados
Regressão Logística
Resultados da Modelagem
Estimativas para os Parâmetros e Distribuições
Estimativas para Razões de Chances e Interpretações
Extensões

Modelos Lineares Generalizados

- As especificações mais comuns para F incluem:

$$F(\mathbf{x}_i'\beta) = \begin{cases} \exp(\mathbf{x}_i'\beta)/[1+\exp(\mathbf{x}_i'\beta)] & \text{logística} \\ \Phi(\mathbf{x}_i'\beta) & \text{probito} \\ 1-\exp[-\exp(\mathbf{x}_i'\beta)] & \text{log log complementar} \end{cases}$$

- Especificando F obtém-se a função de ligação como

$$g(\pi_i) = F^{-1}(\pi_i) = \mathbf{x}_i'\beta.$$

- Com respectivas funções de ligação:

$$g(\pi_i) = \begin{cases} \log\left(\frac{\pi_i}{1-\pi_i}\right), & \text{logística} \\ \Phi^{-1}(\pi_i), & \text{probito} \\ \log(-\log(1-\pi_i)), & \text{log log complementar.} \end{cases}$$

Introdução
Objetivos
Dados Cardiologia
Tabelas de Contingência
Razão de Chances e Risco Relativo
Modelagem Estatística
Modelos Lineares Generalizados
Regressão Logística
Resultados da Modelagem
Estimativas para os Parâmetros e Distribuições
Estimativas para Razões de Chances e Interpretações
Extensões

Modelos Lineares Generalizados

- As especificações mais comuns para F incluem:

$$F(\mathbf{x}_i'\beta) = \begin{cases} \exp(\mathbf{x}_i'\beta)/[1+\exp(\mathbf{x}_i'\beta)] & \text{logística} \\ \Phi(\mathbf{x}_i'\beta) & \text{probito} \\ 1-\exp[-\exp(\mathbf{x}_i'\beta)] & \text{log log complementar} \end{cases}$$

- Especificando F obtém-se a função de ligação como

$$g(\pi_i) = F^{-1}(\pi_i) = \mathbf{x}_i'\beta.$$

- Com respectivas funções de ligação:

$$g(\pi_i) = \begin{cases} \log\left(\frac{\pi_i}{1-\pi_i}\right), & \text{logística} \\ \Phi^{-1}(\pi_i), & \text{probito} \\ \log(-\log(1-\pi_i)), & \text{log log complementar.} \end{cases}$$

Introdução
Objetivos
Dados Cardiologia
Tabelas de Contingência
Razão de Chances e Risco Relativo
Modelagem Estatística
Modelos Lineares Generalizados
Regressão Logística
Resultados da Modelagem
Estimativas para os Parâmetros e Distribuições
Estimativas para Razões de Chances e Interpretações
Extensões

Modelos Lineares Generalizados

- A função de verossimilhança é dada por

$$l(\beta; \mathbf{y}) = \prod_{i=1}^{n} \{F(\mathbf{x}_i'\beta)\}^{y_i} \{1 - F(\mathbf{x}_i'\beta)\}^{(1-y_i)}$$

- Os EMV, $\hat{\beta}$, são obtidos maximizando-se o logaritmo da função de verossimilhança.
- A distribuição assintótica de $\hat{\beta}$ é dada por

$$\hat{\beta} \sim N_{k+1}[\beta,\ I(\hat{\beta})^{-1}],$$

$I(\hat{\beta})$ a matriz de informação de Fisher avaliada em $\hat{\beta}$.

Introdução
Objetivos
Dados Cardiologia
Tabelas de Contingência
Razão de Chances e Risco Relativo
Modelagem Estatística
Modelos Lineares Generalizados
Regressão Logística
Resultados da Modelagem
Estimativas para os Parâmetros e Distribuições
Estimativas para Razões de Chances e Interpretações
Extensões

Modelos Lineares Generalizados

- A função de verossimilhança é dada por

$$l(\beta; \mathbf{y}) = \prod_{i=1}^{n} \{F(\mathbf{x}_i'\beta)\}^{y_i} \{1 - F(\mathbf{x}_i'\beta)\}^{(1-y_i)}$$

- Os EMV, $\hat{\beta}$, são obtidos maximizando-se o logaritmo da função de verossimilhança.
- A distribuição assintótica de $\hat{\beta}$ é dada por

$$\hat{\beta} \sim N_{k+1}[\beta,\ I(\hat{\beta})^{-1}],$$

$I(\hat{\beta})$ a matriz de informação de Fisher avaliada em $\hat{\beta}$.

Introdução
Objetivos
Dados Cardiologia
Tabelas de Contingência
Razão de Chances e Risco Relativo
Modelagem Estatística
Modelos Lineares Generalizados
Regressão Logística
Resultados da Modelagem
Estimativas para os Parâmetros e Distribuições
Estimativas para Razões de Chances e Interpretações
Extensões

Modelos Lineares Generalizados

- A função de verossimilhança é dada por

$$l(\beta; \mathbf{y}) = \prod_{i=1}^{n} \{F(\mathbf{x}_i'\beta)\}^{y_i} \{1 - F(\mathbf{x}_i'\beta)\}^{(1-\, y_i)}$$

- Os EMV, $\hat{\beta}$, são obtidos maximizando-se o logaritmo da função de verossimilhança.
- A distribuição assintótica de $\hat{\beta}$ é dada por

$$\hat{\beta} \sim N_{k+1}[\beta,\ I(\hat{\beta})^{-1}],$$

$I(\hat{\beta})$ a matriz de informação de Fisher avaliada em $\hat{\beta}$.

Introdução
Objetivos
Dados Cardiologia
Tabelas de Contingência
Razão de Chances e Risco Relativo
Modelagem Estatística
Modelos Lineares Generalizados
Regressão Logística
Resultados da Modelagem
Estimativas para os Parâmetros e Distribuições
Estimativas para Razões de Chances e Interpretações
Extensões

Modelo de Regressão Logística

- Modelo de Regressão Logística

$$Y_i|\pi_i \sim Ber(\pi_i), \quad \pi_i = P(Y_i = 1) = F(\mathbf{x}_i'\boldsymbol{\beta}), \quad i = 1, ..., n$$

$$\pi_i = F(\mathbf{x}_i'\boldsymbol{\beta}) = \frac{\exp(\mathbf{x}_i'\boldsymbol{\beta})}{(1 + \exp(\mathbf{x}_i'\boldsymbol{\beta}))}$$

- com função de ligação

$$g(\pi_i) = \log\left(\frac{\pi_i}{1 - \pi_i}\right) = \mathbf{x}_i'\boldsymbol{\beta}$$

$$l(\beta; \mathbf{y}) = \prod_{i=1}^{n} \{F(\mathbf{x}_i'\beta)\}^{y_i} \{1 - F(\mathbf{x}_i'\beta)\}^{(1-\, y_i)}$$

- Se $\widehat{\boldsymbol{\beta}}$ é o Estimador de Máxima Verossimilhança de $\boldsymbol{\beta}$

$$\widehat{\pi}_i = \frac{\exp(\mathbf{x}_i'\widehat{\boldsymbol{\beta}})}{(1 + \exp(\mathbf{x}_i'\widehat{\boldsymbol{\beta}}))}$$

Introdução
Objetivos
Dados Cardiologia
Tabelas de Contingência
Razão de Chances e Risco Relativo
Modelagem Estatística
Modelos Lineares Generalizados
Regressão Logística
Resultados da Modelagem
Estimativas para os Parâmetros e Distribuições
Estimativas para Razões de Chances e Interpretações
Extensões

Modelo de Regressão Logística

- Modelo de Regressão Logística

$$Y_i|\pi_i \sim Ber(\pi_i), \quad \pi_i = P(Y_i = 1) = F(\mathbf{x}_i'\boldsymbol{\beta}), \quad i = 1, ..., n$$

$$\pi_i = F(\mathbf{x}_i'\boldsymbol{\beta}) = \frac{\exp(\mathbf{x}_i'\boldsymbol{\beta})}{(1 + \exp(\mathbf{x}_i'\boldsymbol{\beta}))}$$

- com função de ligação

$$g(\pi_i) = \log\left(\frac{\pi_i}{1 - \pi_i}\right) = \mathbf{x}_i'\boldsymbol{\beta}$$

$$l(\beta; \mathbf{y}) = \prod_{i=1}^{n} \{F(\mathbf{x}_i'\beta)\}^{y_i} \{1 - F(\mathbf{x}_i'\beta)\}^{(1-\, y_i)}$$

- Se $\widehat{\boldsymbol{\beta}}$ é o Estimador de Máxima Verossimilhança de $\boldsymbol{\beta}$

$$\widehat{\pi}_i = \frac{\exp(\mathbf{x}_i'\widehat{\boldsymbol{\beta}})}{(1 + \exp(\mathbf{x}_i'\widehat{\boldsymbol{\beta}}))}$$

Introdução
Objetivos
Dados Cardiologia
Tabelas de Contingência
Razão de Chances e Risco Relativo
Modelagem Estatística
Modelos Lineares Generalizados
Regressão Logística
Resultados da Modelagem
Estimativas para os Parâmetros e Distribuições
Estimativas para Razões de Chances e Interpretações
Extensões

Modelo de Regressão Logística

Introdução
Objetivos
Dados Cardiologia
Tabelas de Contingência
Razão de Chances e Risco Relativo
Modelagem Estatística
Modelos Lineares Generalizados
Regressão Logística
Resultados da Modelagem
Estimativas para os Parâmetros e Distribuições
Estimativas para Razões de Chances e Interpretações

- Modelo de Regressão Logística

$$Y_i|\pi_i \sim Ber(\pi_i), \quad \pi_i = P(Y_i = 1) = F(\mathbf{x}_i'\boldsymbol{\beta}), \quad i = 1, ..., n$$

$$\pi_i = F(\mathbf{x}_i'\boldsymbol{\beta}) = \frac{\exp(\mathbf{x}_i'\boldsymbol{\beta})}{(1 + \exp(\mathbf{x}_i'\boldsymbol{\beta}))}$$

- com função de ligação

$$g(\pi_i) = \log\left(\frac{\pi_i}{1 - \pi_i}\right) = \mathbf{x}_i'\boldsymbol{\beta}$$

$$l(\beta; \mathbf{y}) = \prod_{i=1}^{n} \{F(\mathbf{x}_i'\beta)\}^{y_i} \{1 - F(\mathbf{x}_i'\beta)\}^{(1-\, y_i)}$$

- Se $\widehat{\boldsymbol{\beta}}$ é o Estimador de Máxima Verossimilhança de $\boldsymbol{\beta}$

$$\widehat{\pi}_i = \frac{\exp(\mathbf{x}_i'\widehat{\boldsymbol{\beta}})}{(1 + \exp(\mathbf{x}_i'\widehat{\boldsymbol{\beta}}))}$$

Inferência Bayesiana

- A distribuição a posteriori de $\boldsymbol{\beta}$ é dada por

$$p(\boldsymbol{\beta}|y) = \frac{p(y|\boldsymbol{\beta})p(\boldsymbol{\beta})}{p(y)} \propto p(y|\boldsymbol{\beta})p(\boldsymbol{\beta}),$$

  - $p(y)^{-1}$ constante de proporcionalidade (desempenha importante papel na inferência bayesiana).
  - $p(y) = \int_{\Theta} p(y|\boldsymbol{\beta})p(\boldsymbol{\beta})d\boldsymbol{\beta}$.
  - Na obtenção das distribuições a posteriori surgem integrais analiticamente intratáveis e o uso de métodos numéricos se faz necessário.

Introdução
Objetivos
Dados Cardiologia
Tabelas de Contingência
Razão de Chances e Risco Relativo
Modelagem Estatística
Modelos Lineares Generalizados
Regressão Logística
Resultados da Modelagem
Estimativas para os Parâmetros e Distribuições
Estimativas para Razões de Chances e Interpretações
Extensões

Inferência Bayesiana

- A distribuição a posteriori de $\boldsymbol{\beta}$ é dada por

$$p(\boldsymbol{\beta}|y) = \frac{p(y|\boldsymbol{\beta})p(\boldsymbol{\beta})}{p(y)} \propto p(y|\boldsymbol{\beta})p(\boldsymbol{\beta}),$$

  - $p(y)^{-1}$ constante de proporcionalidade (desempenha importante papel na inferência bayesiana).
  - $p(y) = \int_{\Theta} p(y|\boldsymbol{\beta})p(\boldsymbol{\beta})d\boldsymbol{\beta}$.
  - Na obtenção das distribuições a posteriori surgem integrais analiticamente intratáveis e o uso de métodos numéricos se faz necessário.

Introdução
Objetivos
Dados Cardiologia
Tabelas de Contingência
Razão de Chances e Risco Relativo
Modelagem Estatística
Modelos Lineares Generalizados
Regressão Logística
Resultados da Modelagem
Estimativas para os Parâmetros e Distribuições
Estimativas para Razões de Chances e Interpretações
Extensões

Inferência Bayesiana

- A distribuição a posteriori de $\boldsymbol{\beta}$ é dada por

$$p(\boldsymbol{\beta}|y) = \frac{p(y|\boldsymbol{\beta})p(\boldsymbol{\beta})}{p(y)} \propto p(y|\boldsymbol{\beta})p(\boldsymbol{\beta}),$$

  - $p(y)^{-1}$ constante de proporcionalidade (desempenha importante papel na inferência bayesiana).
  - $p(y) = \int_{\Theta} p(y|\boldsymbol{\beta})p(\boldsymbol{\beta})d\boldsymbol{\beta}$.
  - Na obtenção das distribuições a posteriori surgem integrais analiticamente intratáveis e o uso de métodos numéricos se faz necessário.

Introdução
Objetivos
Dados Cardiologia
Tabelas de Contingência
Razão de Chances e Risco Relativo
Modelagem Estatística
Modelos Lineares Generalizados
Regressão Logística
Resultados da Modelagem
Estimativas para os Parâmetros e Distribuições
Estimativas para Razões de Chances e Interpretações
Extensões

Inferência Bayesiana

- A distribuição a posteriori de $\boldsymbol{\beta}$ é dada por

$$p(\boldsymbol{\beta}|y) = \frac{p(y|\boldsymbol{\beta})p(\boldsymbol{\beta})}{p(y)} \propto p(y|\boldsymbol{\beta})p(\boldsymbol{\beta}),$$

  - $p(y)^{-1}$ constante de proporcionalidade (desempenha importante papel na inferência bayesiana).
  - $p(y) = \int_{\Theta} p(y|\boldsymbol{\beta})p(\boldsymbol{\beta})d\boldsymbol{\beta}$.
  - Na obtenção das distribuições a posteriori surgem integrais analiticamente intratáveis e o uso de métodos numéricos se faz necessário.

Introdução
Objetivos
Dados Cardiologia
Tabelas de Contingência
Razão de Chances e Risco Relativo
Modelagem Estatística
Modelos Lineares Generalizados
Regressão Logística
Resultados da Modelagem
Estimativas para os Parâmetros e Distribuições
Estimativas para Razões de Chances e Interpretações
Extensões

Resultados do Ajuste

Estimativas para os coeficientes do modelo M_4
e respectivos desvios padrão.

	Amostrador de Gibbs		Máxima Verossimilhança	
Parâmetro	Coef.	(desvio padrão)	Coef.	(desvio padrão)
intercepto	-5.787	(1.459)	-5.739	(1.406)
idade	0.047	(0.019)	0.048	(0.018)
sexo	-1.433	(0.443)	-1.406	(0.432)
has	2.407	(1.955)	2.451	(1.945)
iamp	0.193	(0.488)	0.203	(0.477)
fumo	-0.667	(0.399)	-0.612	(0.385)
kient	1.474	(0.199)	1.417	(0.195)
idade x has	-0.054	(0.026)	-0.055	(0.026)
sexo x has	0.617	(0.655)	0.605	(0.638)
iamp x has	1.089	(0.693)	1.051	(0.678)

Introdução
Objetivos
Dados Cardiologia
Tabelas de Contingência
Razão de Chances e Risco Relativo
Modelagem Estatística
Modelos Lineares Generalizados
Regressão Logística
Resultados da Modelagem
Estimativas para os Parâmetros e Distribuições
Estimativas para Razões de Chances e Interpretações
Extensões

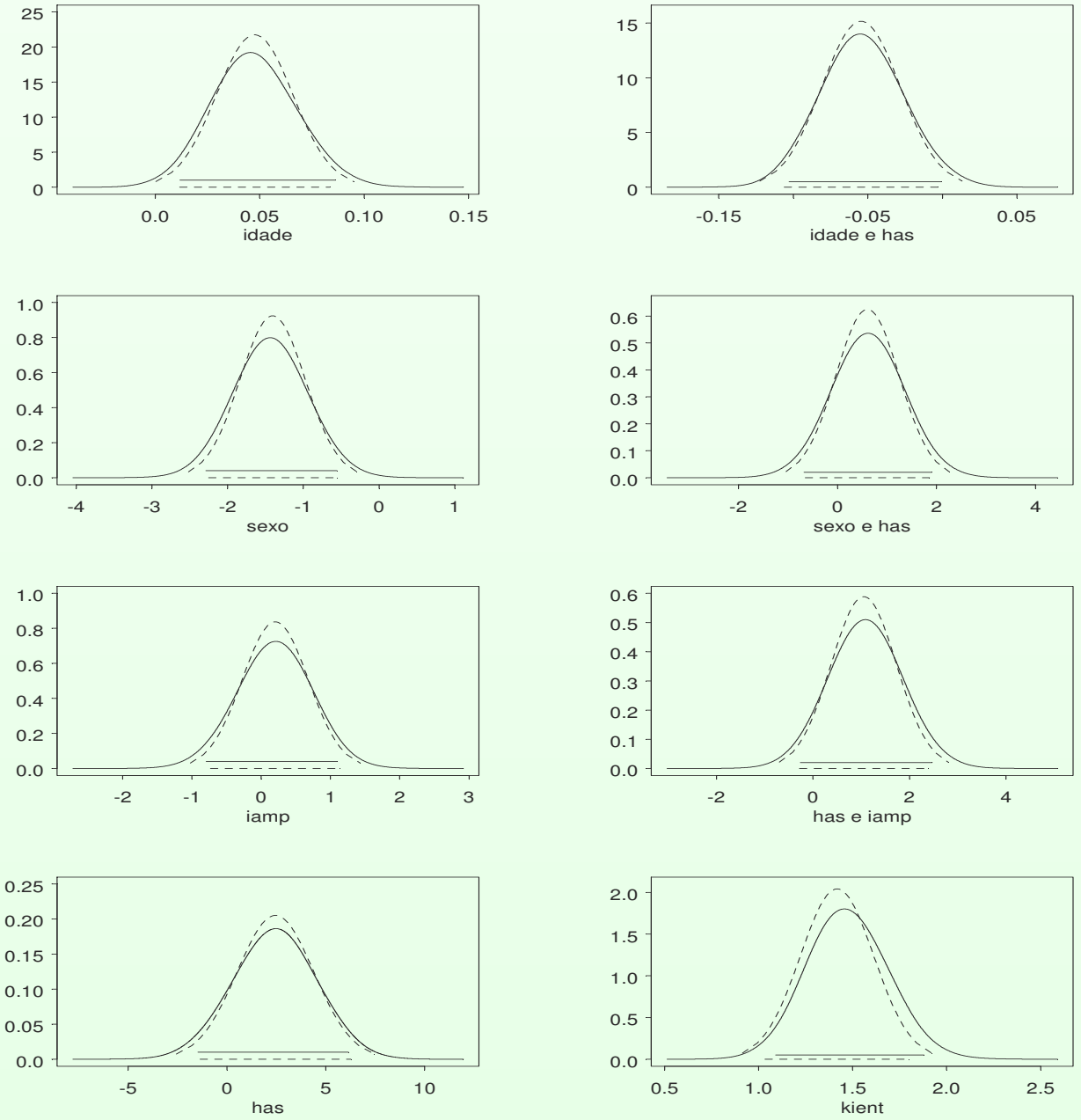


Figura 7.1. Densidades Estimadas para os Coeficientes, sob o modelo M_4, via amostrador de Gibbs (linha contínua) e assintóticas (linha tracejada).

Modelo de Regressão Logística

A Razão de Chances no Modelo de Regressão Logística Simples

variável independente

X

		$x = 1$	$x = 0$
Resp. Y	$y = 1$	$\pi(1) = \frac{e^{\beta_0+\beta_1}}{1+e^{\beta_0+\beta_1}}$	$\pi(0) = \frac{e^{\beta_0}}{1+e^{\beta_0}}$
	$y = 0$	$1-\pi(1) = \frac{1}{1+e^{\beta_0+\beta_1}}$	$1-\pi(0) = \frac{1}{1+e^{\beta_0}}$
	Total	1.0	1.0

Observe que:

- $\pi(1)/(1-\pi(1))$ é a chance do resultado estar presente entre indivíduos com $x = 1$;
- $\pi(0)/(1-\pi(0))$ é a chance do resultado estar presente entre indivíduos com $x = 0$.
- Assim, a razão de chances (odds ratio) é dada por

$$\psi = \frac{\pi(1)/(1-\pi(1))}{\pi(0)/(1-\pi(0))} = \frac{\left(\frac{e^{\beta_0+\beta_1}}{1+e^{\beta_0+\beta_1}}\right)\left(\frac{1}{1+e^{\beta_0}}\right)}{\left(\frac{e^{\beta_0}}{1+e^{\beta_0}}\right)\left(\frac{1}{1+e^{\beta_0+\beta_1}}\right)} = \frac{e^{\beta_0+\beta_1}}{e^{\beta_0}} = e^{\beta_1}$$

Introdução
Objetivos
Dados Cardiologia
Tabelas de Contingência
Razão de Chances e Risco Relativo
Modelagem Estatística
Modelos Lineares Generalizados
Regressão Logística
Resultados da Modelagem
Estimativas para os Parâmetros e Distribuições
Estimativas para Razões de Chances e Interpretações
Extensões

Modelo de Regressão Logística

A Razão de Chances no Modelo de Regressão Logística Simples

variável independente

X

		$x = 1$	$x = 0$
Resp. Y	$y = 1$	$\pi(1) = \frac{e^{\beta_0+\beta_1}}{1+e^{\beta_0+\beta_1}}$	$\pi(0) = \frac{e^{\beta_0}}{1+e^{\beta_0}}$
	$y = 0$	$1-\pi(1) = \frac{1}{1+e^{\beta_0+\beta_1}}$	$1-\pi(0) = \frac{1}{1+e^{\beta_0}}$
	Total	1.0	1.0

Observe que:

- $\pi(1)/(1-\pi(1))$ é a chance do resultado estar presente entre indivíduos com $x = 1$;
- $\pi(0)/(1-\pi(0))$ é a chance do resultado estar presente entre indivíduos com $x = 0$.
- Assim, a razão de chances (odds ratio) é dada por

$$\psi = \frac{\pi(1)/(1-\pi(1))}{\pi(0)/(1-\pi(0))} = \frac{\left(\frac{e^{\beta_0+\beta_1}}{1+e^{\beta_0+\beta_1}}\right)\left(\frac{1}{1+e^{\beta_0}}\right)}{\left(\frac{e^{\beta_0}}{1+e^{\beta_0}}\right)\left(\frac{1}{1+e^{\beta_0+\beta_1}}\right)} = \frac{e^{\beta_0+\beta_1}}{e^{\beta_0}} = e^{\beta_1}$$

Introdução
Objetivos
Dados Cardiologia
Tabelas de Contingência
Razão de Chances e Risco Relativo
Modelagem Estatística
Modelos Lineares Generalizados
Regressão Logística
Resultados da Modelagem
Estimativas para os Parâmetros e Distribuições
Estimativas para Razões de Chances e Interpretações
Extensões

Modelo de Regressão Logística

A Razão de Chances no Modelo de Regressão Logística Simples

variável independente

X

		$x = 1$	$x = 0$
Resp. Y	$y = 1$	$\pi(1) = \frac{e^{\beta_0+\beta_1}}{1+e^{\beta_0+\beta_1}}$	$\pi(0) = \frac{e^{\beta_0}}{1+e^{\beta_0}}$
	$y = 0$	$1-\pi(1) = \frac{1}{1+e^{\beta_0+\beta_1}}$	$1-\pi(0) = \frac{1}{1+e^{\beta_0}}$
	Total	1.0	1.0

Observe que:

- $\pi(1)/(1-\pi(1))$ é a chance do resultado estar presente entre indivíduos com $x = 1$;
- $\pi(0)/(1-\pi(0))$ é a chance do resultado estar presente entre indivíduos com $x = 0$.
- Assim, a razão de chances (odds ratio) é dada por

$$\psi = \frac{\pi(1)/(1-\pi(1))}{\pi(0)/(1-\pi(0))} = \frac{\left(\frac{e^{\beta_0+\beta_1}}{1+e^{\beta_0+\beta_1}}\right)\left(\frac{1}{1+e^{\beta_0}}\right)}{\left(\frac{e^{\beta_0}}{1+e^{\beta_0}}\right)\left(\frac{1}{1+e^{\beta_0+\beta_1}}\right)} = \frac{e^{\beta_0+\beta_1}}{e^{\beta_0}} = e^{\beta_1}$$

Introdução
Objetivos
Dados Cardiologia
Tabelas de Contingência
Razão de Chances e Risco Relativo
Modelagem Estatística
Modelos Lineares Generalizados
Regressão Logística
Resultados da Modelagem
Estimativas para os Parâmetros e Distribuições
Estimativas para Razões de Chances e Interpretações
Extensões

Modelo de Regressão Logística

- Em geral, para uma variável dicotômica x_k, a razão de chances, ajustada para as outras variáveis incluídas no modelo, é dada por

$$\Psi(x_k) = \frac{\exp(\beta_0 + \beta_1 x_1 + ... + \beta_{k-1} x_{k-1} + \beta_k)}{\exp(\beta_0 + \beta_1 x_1 + ... + \beta_{k-1} x_{k-1})} = \exp(\beta_k).$$

- Para variáveis contínuas, a razão de chances, ajustada para as outras variáveis incluídas no modelo, considerando uma mudança de c unidades em x_k, a partir de a, é dada por

$$\Psi(x_k) = \frac{\exp(\beta_0 + \beta_1 x_1 + ... + \beta_{k-1} x_{k-1} + \beta_k (a + c))}{\exp(\beta_0 + \beta_1 x_1 + ... + \beta_{k-1} x_{k-1} + \beta_k a)} = \exp(c\beta_k).$$

- Por causa da assimetria da distribuição amostral de $\hat{\Psi}$, inferências são baseadas na distribuição assintótica de $\hat{\beta}_k = \ln(\hat{\Psi}(x_k))$.
- Intervalos de confiança para razão de chances são obtidos através dos intervalos de confiança para os coeficientes β_k, isto é,

$$\exp[\hat{\beta}_k \pm z_{(1-\alpha/2)} \hat{SE}(\hat{\beta}_k)]$$

onde $\hat{SE}(\hat{\beta}_k)$ é uma estimativa do desvio padrão para o coeficiente β_k.

Introdução
Objetivos
Dados Cardiologia
Tabelas de Contingência
Razão de Chances e Risco Relativo
Modelagem Estatística
Modelos Lineares Generalizados
Regressão Logística
Resultados da Modelagem
Estimativas para os Parâmetros e Distribuições
Estimativas para Razões de Chances e Interpretações
Extensões

Modelo de Regressão Logística

- Em geral, para uma variável dicotômica x_k, a razão de chances, ajustada para as outras variáveis incluídas no modelo, é dada por

$$\Psi(x_k) = \frac{\exp(\beta_0 + \beta_1 x_1 + ... + \beta_{k-1} x_{k-1} + \beta_k)}{\exp(\beta_0 + \beta_1 x_1 + ... + \beta_{k-1} x_{k-1})} = \exp(\beta_k).$$

- Para variáveis contínuas, a razão de chances, ajustada para as outras variáveis incluídas no modelo, considerando uma mudança de c unidades em x_k, a partir de a, é dada por

$$\Psi(x_k) = \frac{\exp(\beta_0 + \beta_1 x_1 + ... + \beta_{k-1} x_{k-1} + \beta_k (a + c))}{\exp(\beta_0 + \beta_1 x_1 + ... + \beta_{k-1} x_{k-1} + \beta_k a)} = \exp(c\beta_k).$$

- Por causa da assimetria da distribuição amostral de $\hat{\Psi}$, inferências são baseadas na distribuição assintótica de $\hat{\beta}_k = \ln(\hat{\Psi}(x_k))$.
- Intervalos de confiança para razão de chances são obtidos através dos intervalos de confiança para os coeficientes β_k, isto é,

$$\exp[\hat{\beta}_k \pm z_{(1-\alpha/2)} \hat{SE}(\hat{\beta}_k)]$$

onde $\hat{SE}(\hat{\beta}_k)$ é uma estimativa do desvio padrão para o coeficiente β_k.

Introdução
Objetivos
Dados Cardiologia
Tabelas de Contingência
Razão de Chances e Risco Relativo
Modelagem Estatística
Modelos Lineares Generalizados
Regressão Logística
Resultados da Modelagem
Estimativas para os Parâmetros e Distribuições
Estimativas para Razões de Chances e Interpretações
Extensões

Modelo de Regressão Logística

- Em geral, para uma variável dicotômica x_k, a razão de chances, ajustada para as outras variáveis incluídas no modelo, é dada por

$$\Psi(x_k) = \frac{\exp(\beta_0 + \beta_1 x_1 + ... + \beta_{k-1} x_{k-1} + \beta_k)}{\exp(\beta_0 + \beta_1 x_1 + ... + \beta_{k-1} x_{k-1})} = \exp(\beta_k).$$

- Para variáveis contínuas, a razão de chances, ajustada para as outras variáveis incluídas no modelo, considerando uma mudança de c unidades em x_k, a partir de a, é dada por

$$\Psi(x_k) = \frac{\exp(\beta_0 + \beta_1 x_1 + ... + \beta_{k-1} x_{k-1} + \beta_k (a + c))}{\exp(\beta_0 + \beta_1 x_1 + ... + \beta_{k-1} x_{k-1} + \beta_k a)} = \exp(c\beta_k).$$

- Por causa da assimetria da distribuição amostral de $\hat{\Psi}$, inferências são baseadas na distribuição assintótica de $\hat{\beta}_k = \ln(\hat{\Psi}\ (x_k))$.
- Intervalos de confiança para razão de chances são obtidos através dos intervalos de confiança para os coeficientes β_k, isto é,

$$\exp[\hat{\beta}_k \pm z_{(1-\alpha/2)}\ \hat{SE}\ (\hat{\beta}_k)]$$

onde $\hat{SE}\ (\hat{\beta}_k)$ é uma estimativa do desvio padrão para o coeficiente β_k.

Introdução
Objetivos
Dados Cardiologia
Tabelas de Contingência
Razão de Chances e Risco Relativo
Modelagem Estatística
Modelos Lineares Generalizados
Regressão Logística
Resultados da Modelagem
Estimativas para os Parâmetros e Distribuições
Estimativas para Razões de Chances e Interpretações
Extensões

Modelo de Regressão Logística

- Em geral, para uma variável dicotômica x_k, a razão de chances, ajustada para as outras variáveis incluídas no modelo, é dada por

$$\Psi(x_k) = \frac{\exp(\beta_0 + \beta_1 x_1 + ... + \beta_{k-1} x_{k-1} + \beta_k)}{\exp(\beta_0 + \beta_1 x_1 + ... + \beta_{k-1} x_{k-1})} = \exp(\beta_k).$$

- Para variáveis contínuas, a razão de chances, ajustada para as outras variáveis incluídas no modelo, considerando uma mudança de c unidades em x_k, a partir de a, é dada por

$$\Psi(x_k) = \frac{\exp(\beta_0 + \beta_1 x_1 + ... + \beta_{k-1} x_{k-1} + \beta_k (a + c))}{\exp(\beta_0 + \beta_1 x_1 + ... + \beta_{k-1} x_{k-1} + \beta_k a)} = \exp(c\beta_k).$$

- Por causa da assimetria da distribuição amostral de $\hat{\Psi}$, inferências são baseadas na distribuição assintótica de $\hat{\beta}_k = \ln(\hat{\Psi}\ (x_k))$.
- Intervalos de confiança para razão de chances são obtidos através dos intervalos de confiança para os coeficientes β_k, isto é,

$$\exp[\hat{\beta}_k \pm z_{(1-\alpha/2)}\ \hat{SE}\ (\hat{\beta}_k)]$$

onde $\hat{SE}\ (\hat{\beta}_k)$ é uma estimativa do desvio padrão para o coeficiente β_k.

Introdução
Objetivos
Dados Cardiologia
Tabelas de Contingência
Razão de Chances e Risco Relativo
Modelagem Estatística
Modelos Lineares Generalizados
Regressão Logística
Resultados da Modelagem
Estimativas para os Parâmetros e Distribuições
Estimativas para Razões de Chances e Interpretações
Extensões

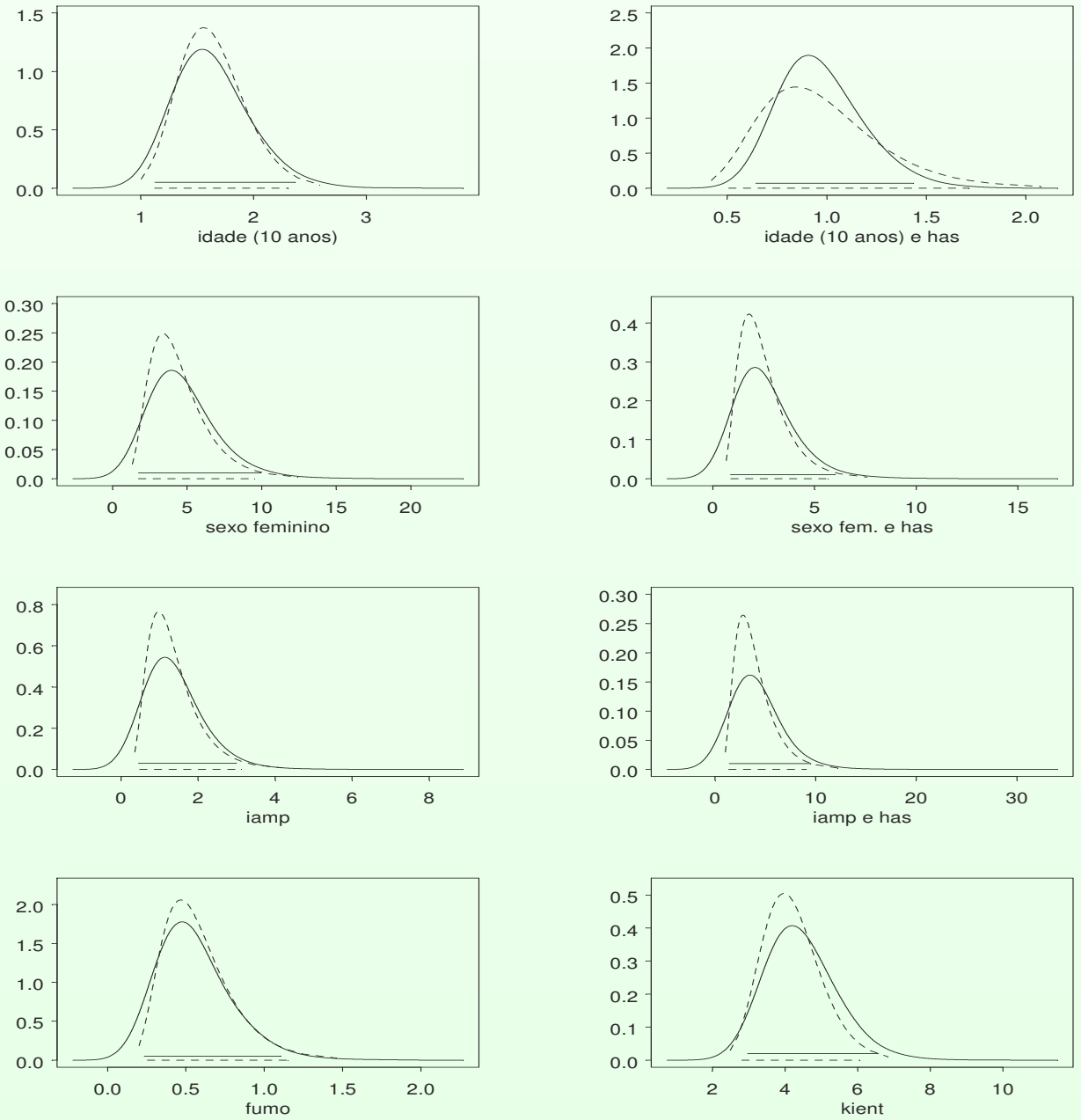


Figura 7.2. Densidades Estimadas para Razões de Chances, sob o modelo M_4, via amostrador de Gibbs (linha contínua) e assintóticas (linha tracejada).

Introdução
Objetivos
Dados Cardiologia
Tabelas de Contingência
Razão de Chances e Risco Relativo
Modelagem Estatística
Modelos Lineares Generalizados
Regressão Logística
Resultados da Modelagem
Estimativas para os Parâmetros e Distribuições
Estimativas para Razões de Chances e Interpretações
Extensões

Interpretação: Razão de Chances

- **idade**, na ausência de hipertensão, é um fator de risco, $\hat{\Psi} = 1.63$ (1.12, 2.37), aumentando em 63 % o risco de óbito a cada 10 anos, enquanto que na presença de hipertensão deixa de ser um fator de risco, $\hat{\Psi} = 0.97$ (0.64, 1.44).
- Para o **sexo** feminino, na ausência de hipertensão, $\hat{\Psi} = 4.69$ (1.75, 9.93), indica que a mortalidade hospitalar é 4.69 vezes mais frequente entre as mulheres, diminuindo para 2.56 (0.87, 6.01) no grupo de mulheres hipertensas.
- Para **iamp**, na presença de hipertensão $\hat{\Psi} = 4.068$ (1.41, 9.50), enquanto que para os não hipertensos este valor diminui para 1.36 (0.45, 3.00).
- Para **has**, considerando idade (10 anos), $\hat{\Psi} = 0.48$ (0.12, 1.35), mostrando que a presença de hipertensão é um fator de proteção.
- Para **fumo**, $\hat{\Psi} = 0.555$ (0.230, 1.10) indica proteção, enquanto que para **kient** $\hat{\Psi} = 4.45$ (2.97, 6.56), indica que o risco de óbito é 4.45 vezes mais frequente a cada unidade aumentada em **kient**.
- Estes resultados mostram que idade, infarto prévio na presença de hipertensão, sexo feminino e killip de entrada são fatores de risco e sexo masculino, hipertensão e fumo representam proteção para óbito hospitalar.

Introdução
Objetivos
Dados Cardiologia
Tabelas de Contingência
Razão de Chances e Risco Relativo
Modelagem Estatística
Modelos Lineares Generalizados
Regressão Logística
Resultados da Modelagem
Estimativas para os Parâmetros e Distribuições
Estimativas para Razões de Chances e Interpretações
Extensões

Interpretação: Razão de Chances

- **idade**, na ausência de hipertensão, é um fator de risco, $\hat{\Psi} = 1.63$ (1.12, 2.37), aumentando em 63 % o risco de óbito a cada 10 anos, enquanto que na presença de hipertensão deixa de ser um fator de risco, $\hat{\Psi} = 0.97$ (0.64, 1.44).
- Para o **sexo** feminino, na ausência de hipertensão, $\hat{\Psi} = 4.69$ (1.75, 9.93), indica que a mortalidade hospitalar é 4.69 vezes mais frequente entre as mulheres, diminuindo para 2.56 (0.87, 6.01) no grupo de mulheres hipertensas.
- Para **iamp**, na presença de hipertensão $\hat{\Psi} = 4.068$ (1.41, 9.50), enquanto que para os não hipertensos este valor diminui para 1.36 (0.45, 3.00).
- Para **has**, considerando idade (10 anos), $\hat{\Psi} = 0.48$ (0.12, 1.35), mostrando que a presença de hipertensão é um fator de proteção.
- Para **fumo**, $\hat{\Psi} = 0.555$ (0.230, 1.10) indica proteção, enquanto que para **kient** $\hat{\Psi} = 4.45$ (2.97, 6.56), indica que o risco de óbito é 4.45 vezes mais frequente a cada unidade aumentada em **kient**.
- Estes resultados mostram que idade, infarto prévio na presença de hipertensão, sexo feminino e killip de entrada são fatores de risco e sexo masculino, hipertensão e fumo representam proteção para óbito hospitalar.

Introdução
Objetivos
Dados Cardiologia
Tabelas de Contingência
Razão de Chances e Risco Relativo
Modelagem Estatística
Modelos Lineares Generalizados
Regressão Logística
Resultados da Modelagem
Estimativas para os Parâmetros e Distribuições
Estimativas para Razões de Chances e Interpretações
Extensões

Interpretação: Razão de Chances

- **idade**, na ausência de hipertensão, é um fator de risco, $\hat{\Psi} = 1.63$ (1.12, 2.37), aumentando em 63 % o risco de óbito a cada 10 anos, enquanto que na presença de hipertensão deixa de ser um fator de risco, $\hat{\Psi} = 0.97$ (0.64, 1.44).
- Para o **sexo** feminino, na ausência de hipertensão, $\hat{\Psi} = 4.69$ (1.75, 9.93), indica que a mortalidade hospitalar é 4.69 vezes mais frequente entre as mulheres, diminuindo para 2.56 (0.87, 6.01) no grupo de mulheres hipertensas.
- Para **iamp**, na presença de hipertensão $\hat{\Psi} = 4.068$ (1.41, 9.50), enquanto que para os não hipertensos este valor diminui para 1.36 (0.45, 3.00).
- Para **has**, considerando idade (10 anos), $\hat{\Psi} = 0.48$ (0.12, 1.35), mostrando que a presença de hipertensão é um fator de proteção.
- Para **fumo**, $\hat{\Psi} = 0.555$ (0.230, 1.10) indica proteção, enquanto que para **kient** $\hat{\Psi} = 4.45$ (2.97, 6.56), indica que o risco de óbito é 4.45 vezes mais frequente a cada unidade aumentada em **kient**.
- Estes resultados mostram que idade, infarto prévio na presença de hipertensão, sexo feminino e killip de entrada são fatores de risco e sexo masculino, hipertensão e fumo representam proteção para óbito hospitalar.

Introdução
Objetivos
Dados Cardiologia
Tabelas de Contingência
Razão de Chances e Risco Relativo
Modelagem Estatística
Modelos Lineares Generalizados
Regressão Logística
Resultados da Modelagem
Estimativas para os Parâmetros e Distribuições
Estimativas para Razões de Chances e Interpretações

Interpretação: Razão de Chances

- **idade**, na ausência de hipertensão, é um fator de risco, $\hat{\Psi} = 1.63$ (1.12, 2.37), aumentando em 63 % o risco de óbito a cada 10 anos, enquanto que na presença de hipertensão deixa de ser um fator de risco, $\hat{\Psi} = 0.97$ (0.64, 1.44).
- Para o **sexo** feminino, na ausência de hipertensão, $\hat{\Psi} = 4.69$ (1.75, 9.93), indica que a mortalidade hospitalar é 4.69 vezes mais frequente entre as mulheres, diminuindo para 2.56 (0.87, 6.01) no grupo de mulheres hipertensas.
- Para **iamp**, na presença de hipertensão $\hat{\Psi} = 4.068$ (1.41, 9.50), enquanto que para os não hipertensos este valor diminui para 1.36 (0.45, 3.00).
- Para **has**, considerando idade (10 anos), $\hat{\Psi} = 0.48$ (0.12, 1.35), mostrando que a presença de hipertensão é um fator de proteção.
- Para **fumo**, $\hat{\Psi} = 0.555$ (0.230, 1.10) indica proteção, enquanto que para **kient** $\hat{\Psi} = 4.45$ (2.97, 6.56), indica que o risco de óbito é 4.45 vezes mais frequente a cada unidade aumentada em **kient**.
- Estes resultados mostram que idade, infarto prévio na presença de hipertensão, sexo feminino e killip de entrada são fatores de risco e sexo masculino, hipertensão e fumo representam proteção para óbito hospitalar.

Introdução
Objetivos
Dados Cardiologia
Tabelas de Contingência
Razão de Chances e Risco Relativo
Modelagem Estatística
Modelos Lineares Generalizados
Regressão Logística
Resultados da Modelagem
Estimativas para os Parâmetros e Distribuições
Estimativas para Razões de Chances e Interpretações
Extensões

Interpretação: Razão de Chances

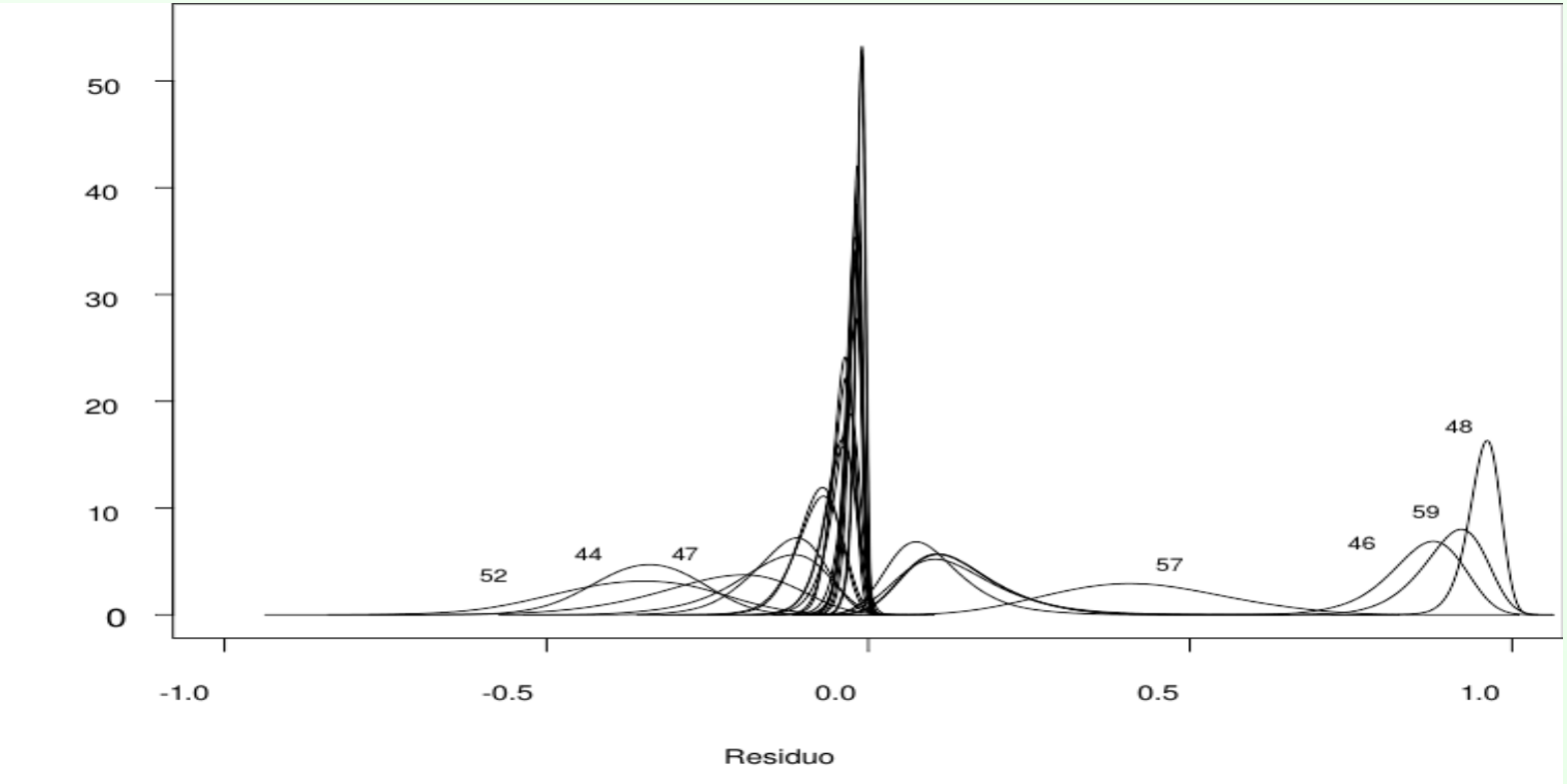
- **idade**, na ausência de hipertensão, é um fator de risco, $\hat{\Psi} = 1.63$ (1.12, 2.37), aumentando em 63 % o risco de óbito a cada 10 anos, enquanto que na presença de hipertensão deixa de ser um fator de risco, $\hat{\Psi} = 0.97$ (0.64, 1.44).
- Para o **sexo** feminino, na ausência de hipertensão, $\hat{\Psi} = 4.69$ (1.75, 9.93), indica que a mortalidade hospitalar é 4.69 vezes mais frequente entre as mulheres, diminuindo para 2.56 (0.87, 6.01) no grupo de mulheres hipertensas.
- Para **iamp**, na presença de hipertensão $\hat{\Psi} = 4.068$ (1.41, 9.50), enquanto que para os não hipertensos este valor diminui para 1.36 (0.45, 3.00).
- Para **has**, considerando idade (10 anos), $\hat{\Psi} = 0.48$ (0.12, 1.35), mostrando que a presença de hipertensão é um fator de proteção.
- Para **fumo**, $\hat{\Psi} = 0.555$ (0.230, 1.10) indica proteção, enquanto que para **kient** $\hat{\Psi} = 4.45$ (2.97, 6.56), indica que o risco de óbito é 4.45 vezes mais frequente a cada unidade aumentada em **kient**.
- Estes resultados mostram que idade, infarto prévio na presença de hipertensão, sexo feminino e killip de entrada são fatores de risco e sexo masculino, hipertensão e fumo representam proteção para óbito hospitalar.

Introdução
Objetivos
Dados Cardiologia
Tabelas de Contingência
Razão de Chances e Risco Relativo
Modelagem Estatística
Modelos Lineares Generalizados
Regressão Logística
Resultados da Modelagem
Estimativas para os Parâmetros e Distribuições
Estimativas para Razões de Chances e Interpretações
Extensões

Interpretação: Razão de Chances

- **idade**, na ausência de hipertensão, é um fator de risco, $\hat{\Psi}= 1.63$ (1.12, 2.37), aumentando em 63 % o risco de óbito a cada 10 anos, enquanto que na presença de hipertensão deixa de ser um fator de risco, $\hat{\Psi}= 0.97$ (0.64, 1.44).
- Para o **sexo** feminino, na ausência de hipertensão, $\hat{\Psi}= 4.69$ (1.75, 9.93), indica que a mortalidade hospitalar é 4.69 vezes mais frequente entre as mulheres, diminuindo para 2.56 (0.87, 6.01) no grupo de mulheres hipertensas.
- Para **iamp**, na presença de hipertensão $\hat{\Psi}= 4.068$ (1.41, 9.50), enquanto que para os não hipertensos este valor diminui para 1.36 (0.45, 3.00).
- Para **has**, considerando idade (10 anos), $\hat{\Psi} = 0.48$ (0.12, 1.35), mostrando que a presença de hipertensão é um fator de proteção.
- Para **fumo**, $\hat{\Psi} = 0.555$ (0.230, 1.10) indica proteção, enquanto que para **kient** $\hat{\Psi} = 4.45$ (2.97, 6.56), indica que o risco de óbito é 4.45 vezes mais frequente a cada unidade aumentada em **kient**.
- Estes resultados mostram que idade, infarto prévio na presença de hipertensão, sexo feminino e killip de entrada são fatores de risco e sexo masculino, hipertensão e fumo representam proteção para óbito hospitalar.

Introdução
Objetivos
Dados Cardiologia
Tabelas de Contingência
Razão de Chances e Risco Relativo
Modelagem Estatística
Modelos Lineares Generalizados
Regressão Logística
Resultados da Modelagem
Estimativas para os Parâmetros e Distribuições
Estimativas para Razões de Chances e Interpretações
Extensões



Densidades a Posteriori para os Residuos $r_i = y_i - \pi_i$, $i = 1, 2, \cdots, 60$.

Introdução
Objetivos
Dados Cardiologia
Tabelas de Contingência
Razão de Chances e Risco Relativo
Modelagem Estatística
Modelos Lineares Generalizados
Regressão Logística
Resultados da Modelagem
Estimativas para os Parâmetros e Distribuições
Estimativas para Razões de Chances e Interpretações
Extensões

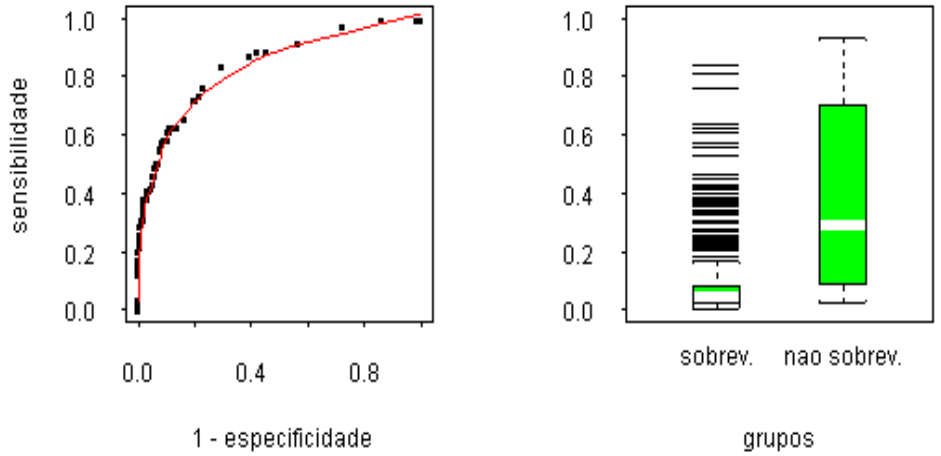


Figura 5.3. Curva ROC (c-index = 0.84 e desvio padrão = 0.02 obtidos via Jackknife) e Boxplots, para as distribuições das probabilidades de óbito estimadas em cada grupo.

Introdução
Objetivos
Dados Cardiologia
Tabelas de Contingência
Razão de Chances e Risco Relativo
Modelagem Estatística
Modelos Lineares Generalizados
Regressão Logística
Resultados da Modelagem
Estimativas para os Parâmetros e Distribuições
Estimativas para Razões de Chances e Interpretações

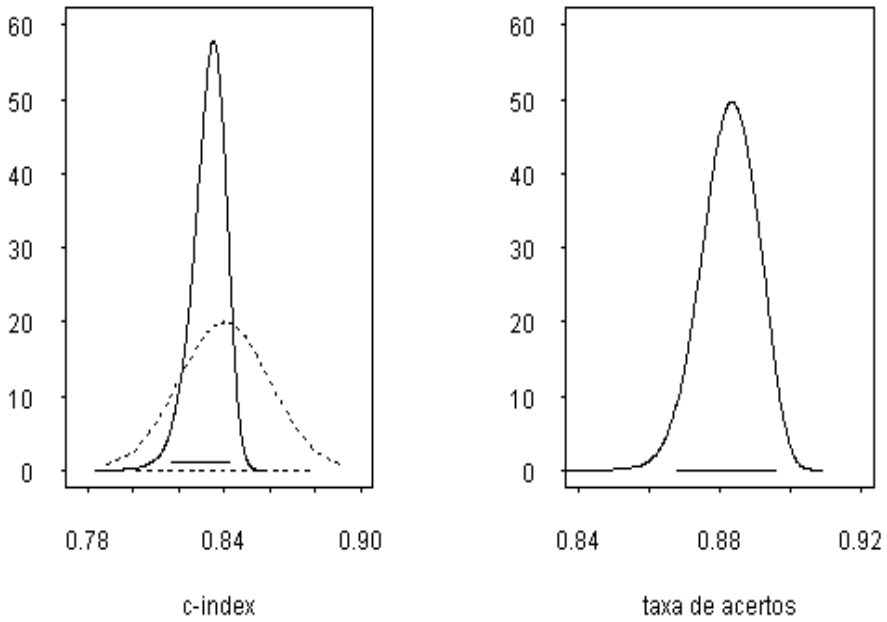


Figura 5.4. Densidades estimadas via Gibbs Sampler (linha contínua) para o c-index e para taxa de acertos, e via Jackknife (linha tracejada) para o c-index.

Introdução
Objetivos
Dados Cardiologia
Tabelas de Contingência
Razão de Chances e Risco Relativo
Modelagem Estatística
Modelos Lineares Generalizados
Regressão Logística
Resultados da Modelagem
Estimativas para os Parâmetros e Distribuições
Estimativas para Razões de Chances e Interpretações
Extensões

Modelo de Regressão Logística com Efeito Aleatório

O modelo é especificado como

$$\begin{aligned} Y_i|\pi_i &\sim Ber(\pi_i), \\ \pi_i &= F(\mathbf{x}_i'\beta + \gamma_i), \qquad i = 1, ..., n \end{aligned}$$

- onde $Y_i = 1$ (0) se a resposta de interesse é (não) observada;
- π_i é a probabilidade de $Y_i = 1$;
- $\beta = (\beta_0, \beta_1, \beta_2, ..., \beta_k)'$ vetor de parâmetros desconhecidos;
- $\mathbf{x}_i = (1, x_{i1}, x_{i2}, ..., x_{ik})'$ o vetor de covariáveis conhecidas;
- F função distribuição acumulada, por exemplo, da logistica, dada por

$$\pi_i = \frac{\exp(\mathbf{x}_i'\beta + \gamma_i)}{[1 + \exp(\mathbf{x}_i'\beta + \gamma_i)]}.$$

Introdução
Objetivos
Dados Cardiologia
Tabelas de Contingência
Razão de Chances e Risco Relativo
Modelagem Estatística
Modelos Lineares Generalizados
Regressão Logística
Resultados da Modelagem
Estimativas para os Parâmetros e Distribuições
Estimativas para Razões de Chances e Interpretações
Extensões

Distribuições a Priori Alternativas para o Efeito Aleatório

- Modelo Normal (M_1):

$$\gamma_i|\sigma^2 \sim N(0, \sigma^2);$$

- Mistura de Modelos Normais para o parâmetro de escala (Discreta) (M_2):

$$\gamma_i|k_i, \sigma^2 \sim N(0, [(1-k_i)\sigma^2 + ck_i\sigma^2]); \;\; k_i|p \sim Ber(p);$$

As quantidades c e p são apropriadas constantes conhecidas.

- Mistura de modelos Normais para o parâmetro de escala (t-Student) (M_3):

$$\gamma_i|\lambda_i, \sigma^2 \sim N(0, \lambda_i\sigma^2), \;\; \lambda_i^{-1}|\nu \sim Ga(\frac{\nu}{2}, \frac{\nu}{2})$$

com alguma priori para ν; A priori de Jeffreys é fácil de implementar (Fonseca, Ferreira and Migon, 2005).

Introdução
Objetivos
Dados Cardiologia
Tabelas de Contingência
Razão de Chances e Risco Relativo
Modelagem Estatística
Modelos Lineares Generalizados
Regressão Logística
Resultados da Modelagem
Estimativas para os Parâmetros e Distribuições
Estimativas para Razões de Chances e Interpretações
Extensões

Distribuições a Priori Alternativas para o Efeito Aleatório

- Modelo Normal (M_1):

$$\gamma_i|\sigma^2 \sim N(0, \sigma^2);$$

- Mistura de Modelos Normais para o parâmetro de escala (Discreta) (M_2):

$$\gamma_i|k_i, \sigma^2 \sim N(0, [(1-k_i)\sigma^2 + ck_i\sigma^2]); \;\; k_i|p \sim Ber(p);$$

As quantidades c e p são apropriadas constantes conhecidas.

- Mistura de modelos Normais para o parâmetro de escala (t-Student) (M_3):

$$\gamma_i|\lambda_i, \sigma^2 \sim N(0, \lambda_i\sigma^2), \;\; \lambda_i^{-1}|\nu \sim Ga(\frac{\nu}{2}, \frac{\nu}{2})$$

com alguma priori para ν; A priori de Jeffreys é fácil de implementar (Fonseca, Ferreira and Migon, 2005).

Introdução
Objetivos
Dados Cardiologia
Tabelas de Contingência
Razão de Chances e Risco Relativo
Modelagem Estatística
Modelos Lineares Generalizados
Regressão Logística
Resultados da Modelagem
Estimativas para os Parâmetros e Distribuições
Estimativas para Razões de Chances e Interpretações
Extensões

Distribuições a Priori Alternativas para o Efeito Aleatório

- Modelo Normal (M_1):

$$\gamma_i|\sigma^2 \sim N(0, \sigma^2);$$

- Mistura de Modelos Normais para o parâmetro de escala (Discreta) (M_2):

$$\gamma_i|k_i, \sigma^2 \sim N(0, [(1-k_i)\sigma^2 + ck_i\sigma^2]); \;\; k_i|p \sim Ber(p);$$

  As quantidades c e p são apropriadas constantes conhecidas.
- Mistura de modelos Normais para o parâmetro de escala (t-Student) (M_3):

$$\gamma_i|\lambda_i, \sigma^2 \sim N(0, \lambda_i\sigma^2), \;\; \lambda_i^{-1}|\nu \sim Ga(\frac{\nu}{2}, \frac{\nu}{2})$$

  com alguma priori para ν; A priori de Jeffreys é fácil de implementar (Fonseca, Ferreira and Migon, 2005).

Introdução
Objetivos
Dados Cardiologia
Tabelas de Contingência
Razão de Chances e Risco Relativo
Modelagem Estatística
Modelos Lineares Generalizados
Regressão Logística
Resultados da Modelagem
Estimativas para os Parâmetros e Distribuições
Estimativas para Razões de Chances e Interpretações
Extensões

Distribuições Alternativas a Priori para σ_2

- Inversa Gama

$$\sigma^{-2} \sim Ga(a, b)$$

com a e b constantes apropriadas;

- Uniforme

$$\sigma \sim Un(0, U)$$

usadas em todos os casos (Gelman, 2006).

Introdução
Objetivos
Dados Cardiologia
Tabelas de Contingência
Razão de Chances e Risco Relativo
Modelagem Estatística
Modelos Lineares Generalizados
Regressão Logística
Resultados da Modelagem
Estimativas para os Parâmetros e Distribuições
Estimativas para Razões de Chances e Interpretações
Extensões

Distribuições Alternativas a Priori para σ_2

- Inversa Gama

$$\sigma^{-2} \sim Ga(a, b)$$

com a e b constantes apropriadas;

- Uniforme

$$\sigma \sim Un(0, U)$$

usadas em todos os casos (Gelman, 2006).

Introdução
Objetivos
Dados Cardiologia
Tabelas de Contingência
Razão de Chances e Risco Relativo
Modelagem Estatística
Modelos Lineares Generalizados
Regressão Logística
Resultados da Modelagem
Estimativas para os Parâmetros e Distribuições
Estimativas para Razões de Chances e Interpretações
Extensões

Distribuição a Posteriori

A distribuição a posteriori conjunta para β, γ and θ_j é dada por

$$p(\beta, \gamma, \theta_j|\mathbf{y}) \propto l(\beta, \gamma; \mathbf{y}) p(\beta) p(\gamma|\theta_j) p(\theta_j), \quad j = 1, 2, 3$$

onde

- $l(\beta, \gamma; \mathbf{y}) = \prod_{i=1}^{n} \left[F(\mathbf{x}_i'\beta + \gamma_i)\right]^{y_i} \left[1 - F(\mathbf{x}_i'\beta + \gamma_i)\right]^{1-y_i}$;
- $\mathbf{y} = (y_1, y_2, ..., y_n)'$ vetor dos dados observados;
- $\pi = (\pi_1, \pi_2, ..., \pi_n)'$;
- $\theta_1 = \sigma^2; \quad \theta_2 = (\mathbf{k}, \sigma^2); \quad \theta_3 = (\lambda, \sigma^2, \nu)$;
- $\mathbf{k} = (k_1, ..., k_n)'$ and $\lambda = (\lambda_1, ..., \lambda_n)'$.

Métodos MCMC (Monte Carlo Markov Chain) foram usados para a obtenção das distriuições marginais a posteriori. (Gamerman & Lopes, 2006).

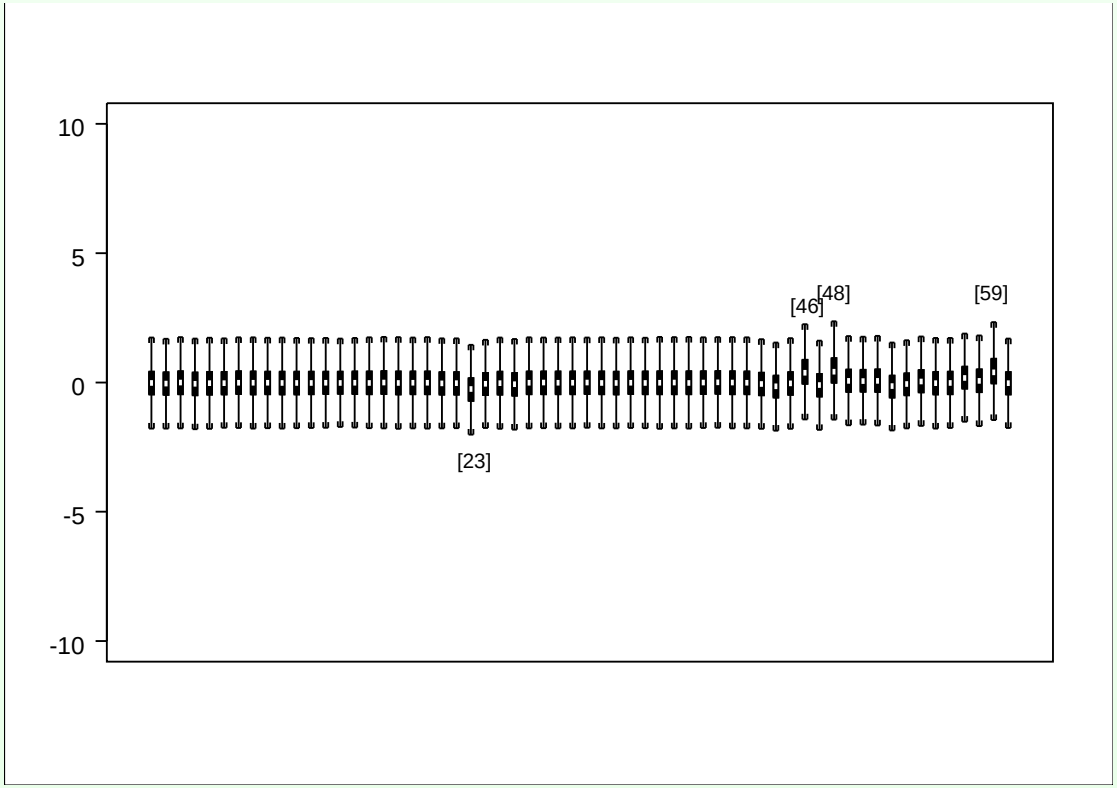
Introdução
Objetivos
Dados Cardiologia
Tabelas de Contingência
Razão de Chances e Risco Relativo
Modelagem Estatística
Modelos Lineares Generalizados
Regressão Logística
Resultados da Modelagem
Estimativas para os Parâmetros e Distribuições
Estimativas para Razões de Chances e Interpretações
Extensões

Diagnósticos de Classificação

Diagnósticos de Classificação para os Modelos M_0, M_1, M_2 $(c = 3)$ and M_3 com Intervalos de Credibilidade de 95% (ponto de corte = 0.55).

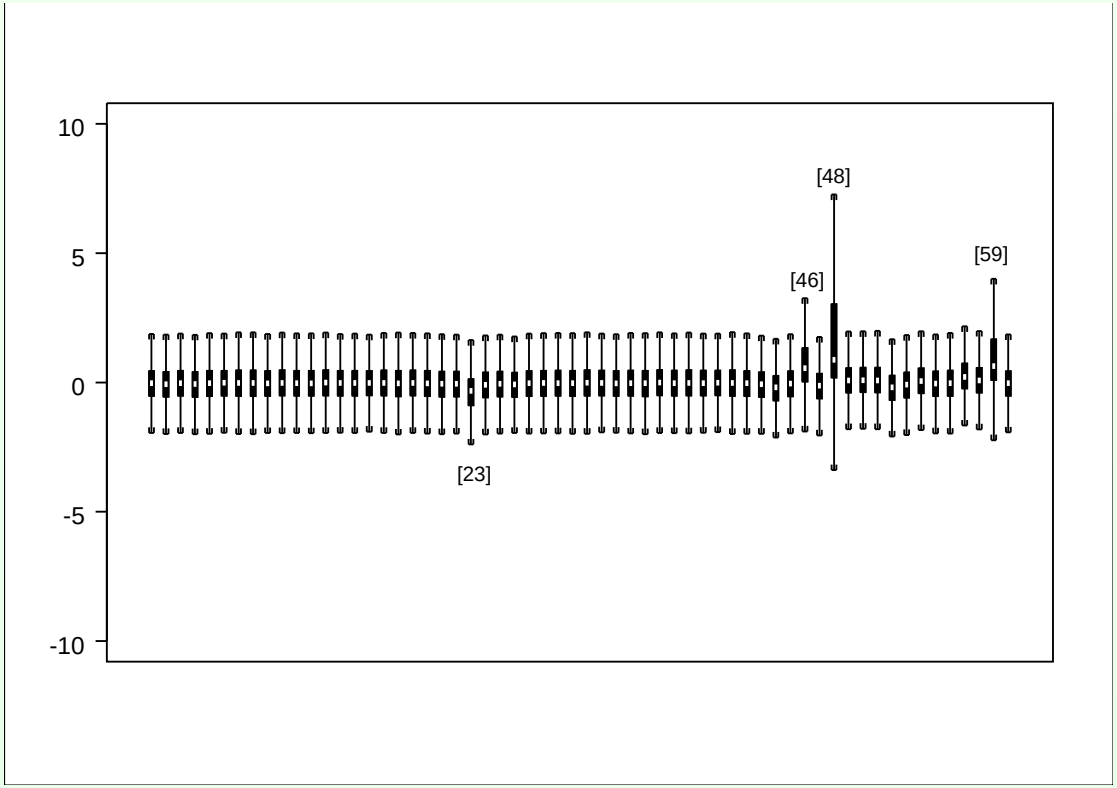
modelo	c-index	taxa de acerto
M_0	0.833 (0.816; 0.843)	0.891 (0.881; 0.897)
M_1	0.919 (0.865; 0.953)	0.914 (0.892; 0.932)
M_2	0.928 (0.867; 0.960)	0.918 (0.894; 0.938)
M_3	0.935 (0.882; 0.966)	0.921 (0.897; 0.941)

Introdução
Objetivos
Dados Cardiologia
Tabelas de Contingência
Razão de Chances e Risco Relativo
Modelagem Estatística
Modelos Lineares Generalizados
Regressão Logística
Resultados da Modelagem
Estimativas para os Parâmetros e Distribuições
Estimativas para Razões de Chances e Interpretações
Extensões



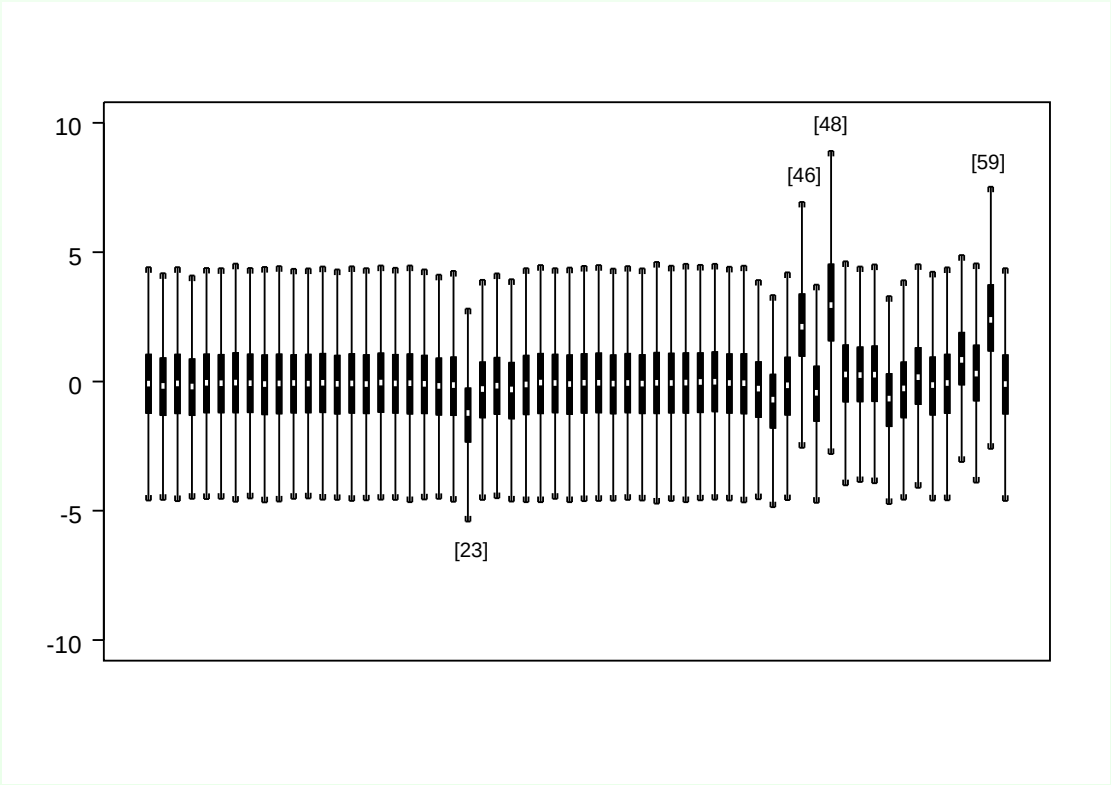
Box-plots para as distribuições dos efeitos aleatórios, modelo M_1, (primeiros 60 pacientes).

Introdução
Objetivos
Dados Cardiologia
Tabelas de Contingência
Razão de Chances e Risco Relativo
Modelagem Estatística
Modelos Lineares Generalizados
Regressão Logística
Resultados da Modelagem
Estimativas para os Parâmetros e Distribuições
Estimativas para Razões de Chances e Interpretações
Extensões



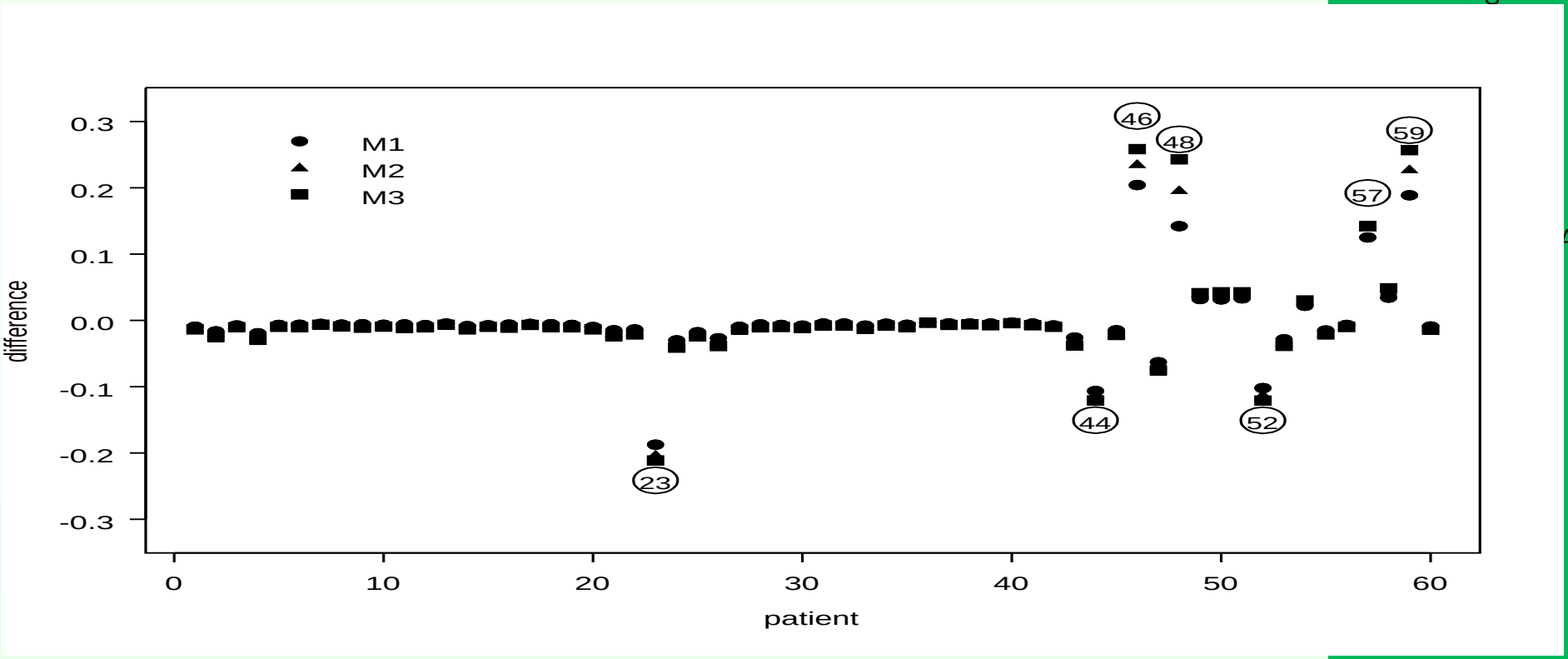
Box-plots para as distribuições dos efeitos aleatórios, modelo M_2, (primeiros 60 pacientes).

Introdução
Objetivos
Dados Cardiologia
Tabelas de Contingência
Razão de Chances e Risco Relativo
Modelagem Estatística
Modelos Lineares Generalizados
Regressão Logística
Resultados da Modelagem
Estimativas para os Parâmetros e Distribuições
Estimativas para Razões de Chances e Interpretações
Extensões



Box-plots para as distribuições dos efeitos aleatórios, modelo M_3, (primeiros 60 pacientes).

Introdução
Objetivos
Dados Cardiologia
Tabelas de Contingência
Razão de Chances e Risco Relativo
Modelagem Estatística
Modelos Lineares Generalizados
Regressão Logística
Resultados da Modelagem
Estimativas para os Parâmetros e Distribuições
Estimativas para Razões de Chances e Interpretações
Extensões



Comparação dos resultados para as probabilidades médias de óbito hospitalar a posteriori considerando os modelos M_1, M_2 and M_3 em relação ao modelo M_0 (primeiros 60 pacientes).

Modelagem Estatística
Modelos Lineares Generalizados
Regressão Logística
Resultados da Modelagem
Estimativas para os Parâmetros e Distribuições
Estimativas para Razões de Chances e Interpretações

Table 3. Diagnósticos de Classificação para M_0 incluindo ou não os candidatos a outlier (ponto de corte = 0.55).

n	c-index	taxa de classificação correta
546	0.833 (0.817;0.843)	0.891 (0.881; 0.897)
519	0.965 (0.958;0.969)	0.939 (0.930; 0.946)
506	0.988 (0.985;0.989)	0.966 (0.958; 0.972)

O Amostrador de Gibbs foi implementado facilmente através do software Win**BUGS** (Bayesian Inference Using Gibbs Sampler) [Spiegelhalter et al., 1996].

Introdução
Objetivos
Dados Cardiologia
Tabelas de Contingência
Razão de Chances e Risco Relativo
Modelagem Estatística
Modelos Lineares Generalizados
Regressão Logística
Resultados da Modelagem
Estimativas para os Parâmetros e Distribuições
Estimativas para Razões de Chances e Interpretações
Extensões

Agradecimentos e Principais Referências

Agradecimentos:

À Comissão organizadora do Evento.

Principais Referências

- Gamerman, D. e Lopes, H. F. (2006) *Markov Chain Monte Carlo: Stochastic Simulation for Baysian Inference*. London: Chapman & Hall/CRC, 2nd. ed.
- Migon, H. S., Gamerman, D. e Louzada, F. (2014) it Statistical Inference: an Integrated Approach. 2. ed. Boca Raton FL: Chapman Hall, 2014. v. 1. 367p.
- Migon, H. S., Souza, A. D. P, Schmidt, A. M. (2008) *Modelos Hierárquicos e Aplicações*. Texto para Minicurso do 18^0_- SINAPE.
  site do livro: www.dme.ufrj.br/ mhierarquico
- Souza, A. D. P. ; MIGON, H. S. (2010) *Bayesian outlier analysis in binary regression*. Journal of Applied Statistics, v. 37, p. 1355-1368.

Introdução
Objetivos
Dados Cardiologia
Tabelas de Contingência
Razão de Chances e Risco Relativo
Modelagem Estatística
Modelos Lineares Generalizados
Regressão Logística
Resultados da Modelagem
Estimativas para os Parâmetros e Distribuições
Estimativas para Razões de Chances e Interpretações

Ganho de Peso em Gestantes

Análise do Ganho de Peso em Gestantes
Motivando Regressão Hierárquica
Modelos para a Análise do Ganho de Peso em Gestantes
Modelo hierárquico de dois estágios
Modelo hierárquico de três estágios
Modelo de Crescimento Linear

Dados sobre Ganho de Peso em Gestantes

Deficiências nutricionais interferem na qualidade da gravidez e repercutem no peso das crianças ao nascer.

⇓

Ministério da Saúde adotou um Instrumento para avaliação do estado nutricional baseado numa curva de monitorização do ganho de peso.

⇓

Curva estabelecida em dados de gestantes não brasileiras.

⇓

Problema

Nutricionistas acreditam que este Instrumento exija um ganho de peso acima do necessário para um bom resultado na gestação.

Análise do Ganho de Peso em Gestantes
Motivando Regressão Hierárquica
Modelos para a Análise do Ganho de Peso em Gestantes
Modelo hierárquico de dois estágios
Modelo hierárquico de três estágios
Modelo de Crescimento Linear

Dados sobre Ganho de Peso em Gestantes

Deficiências nutricionais interferem na qualidade da gravidez e repercutem no peso das crianças ao nascer.

$\Downarrow$

Ministério da Saúde adotou um Instrumento para avaliação do estado nutricional baseado numa curva de monitorização do ganho de peso.

$\Downarrow$

Curva estabelecida em dados de gestantes não brasileiras.

$\Downarrow$

Problema

Nutricionistas acreditam que este Instrumento exija um ganho de peso acima do necessário para um bom resultado na gestação.

Análise do Ganho de Peso em Gestantes
Motivando Regressão Hierárquica
Modelos para a Análise do Ganho de Peso em Gestantes
Modelo hierárquico de dois estágios
Modelo hierárquico de três estágios
Modelo de Crescimento Linear

Dados sobre Ganho de Peso em Gestantes

Deficiências nutricionais interferem na qualidade da gravidez e repercutem no peso das crianças ao nascer.

$\Downarrow$

Ministério da Saúde adotou um Instrumento para avaliação do estado nutricional baseado numa curva de monitorização do ganho de peso.

$\Downarrow$

Curva estabelecida em dados de gestantes não brasileiras.

$\Downarrow$

Problema

Nutricionistas acreditam que este Instrumento exija um ganho de peso acima do necessário para um bom resultado na gestação.

Análise do Ganho de Peso em Gestantes
Motivando Regressão Hierárquica
Modelos para a Análise do Ganho de Peso em Gestantes
Modelo hierárquico de dois estágios
Modelo hierárquico de três estágios
Modelo de Crescimento Linear

Dados sobre Ganho de Peso em Gestantes

- Referem-se a 68 gestantes (inicialmente 120)
  Inst. Puericultura e Pediatria Martagão Gesteira (IPPMG - UFRJ)
  Junho 90 a Janeiro 93

- Principais Características:
  - Primigestas;
  - 20 a 30 anos;
  - não fumantes;
  - não classificadas como gestantes de alto risco;
  - com menos de 15 semanas de gestação;

- Classificadas em Diferentes Grupos (Rosso, 1985):
  - Desnutridas (19 gest.; pré $< 90\%$ peso padrão);
  - Normais (32 gest.; entre 90 e $< 110\%$ peso padrão);
  - Obesas (17 gest.; pré $> 110\%$ peso padrão).

Análise do Ganho de Peso em Gestantes
Motivando Regressão Hierárquica
Modelos para a Análise do Ganho de Peso em Gestantes
Modelo hierárquico de dois estágios
Modelo hierárquico de três estágios
Modelo de Crescimento Linear

Dados sobre Ganho de Peso em Gestantes

- Referem-se a 68 gestantes (inicialmente 120)
  Inst. Puericultura e Pediatria Martagão Gesteira (IPPMG - UFRJ)
  Junho 90 a Janeiro 93
- Principais Características:
  - Primigestas;
  - 20 a 30 anos;
  - não fumantes;
  - não classificadas como gestantes de alto risco;
  - com menos de 15 semanas de gestação;
- Classificadas em Diferentes Grupos (Rosso, 1985):
  - Desnutridas (19 gest.; pré $< 90\%$ peso padrão);
  - Normais (32 gest.; entre 90 e $< 110\%$ peso padrão);
  - Obesas (17 gest.; pré $> 110\%$ peso padrão).

Análise do Ganho de Peso em Gestantes
Motivando Regressão Hierárquica
Modelos para a Análise do Ganho de Peso em Gestantes
Modelo hierárquico de dois estágios
Modelo hierárquico de três estágios
Modelo de Crescimento Linear

Dados sobre Ganho de Peso em Gestantes

- Referem-se a 68 gestantes (inicialmente 120)
  Inst. Puericultura e Pediatria Martagão Gesteira (IPPMG - UFRJ)
  Junho 90 a Janeiro 93
- Principais Características:
  - Primigestas;
  - 20 a 30 anos;
  - não fumantes;
  - não classificadas como gestantes de alto risco;
  - com menos de 15 semanas de gestação;
- Classificadas em Diferentes Grupos (Rosso, 1985):
  - Desnutridas (19 gest.; pré $< 90\%$ peso padrão);
  - Normais (32 gest.; entre 90 e $< 110\%$ peso padrão);
  - Obesas (17 gest.; pré $> 110\%$ peso padrão).

Análise do Ganho de Peso em Gestantes
Motivando Regressão Hierárquica
Modelos para a Análise do Ganho de Peso em Gestantes
Modelo hierárquico de dois estágios
Modelo hierárquico de três estágios
Modelo de Crescimento Linear

Dados sobre Ganho de Peso em Gestantes

Tipo de Observação: Dados Longitudinais

- irregulares (intervalos entre medidas consecutivas não constantes ao longo do tempo);
- não balanceados (observações feitas em diferentes instantes de tempo);
- incompletos (observações perdidas).

Objetivos:

- Estudar o comportamento do ganho de peso ao longo da gestação, para avaliar a adequabilidade do instrumento;
- Verificar se existem diferenças no comportamento do ganho de peso, de acordo com o estado nutricional na concepção.

Dados sobre Ganho de Peso em Gestantes

Análise do Ganho de Peso em Gestantes
Motivando Regressão Hierárquica
Modelos para a Análise do Ganho de Peso em Gestantes
Modelo hierárquico de dois estágios
Modelo hierárquico de três estágios
Modelo de Crescimento Linear

Tipo de Observação: Dados Longitudinais

- irregulares (intervalos entre medidas consecutivas não constantes ao longo do tempo);
- não balanceados (observações feitas em diferentes instantes de tempo);
- incompletos (observações perdidas).

Objetivos:

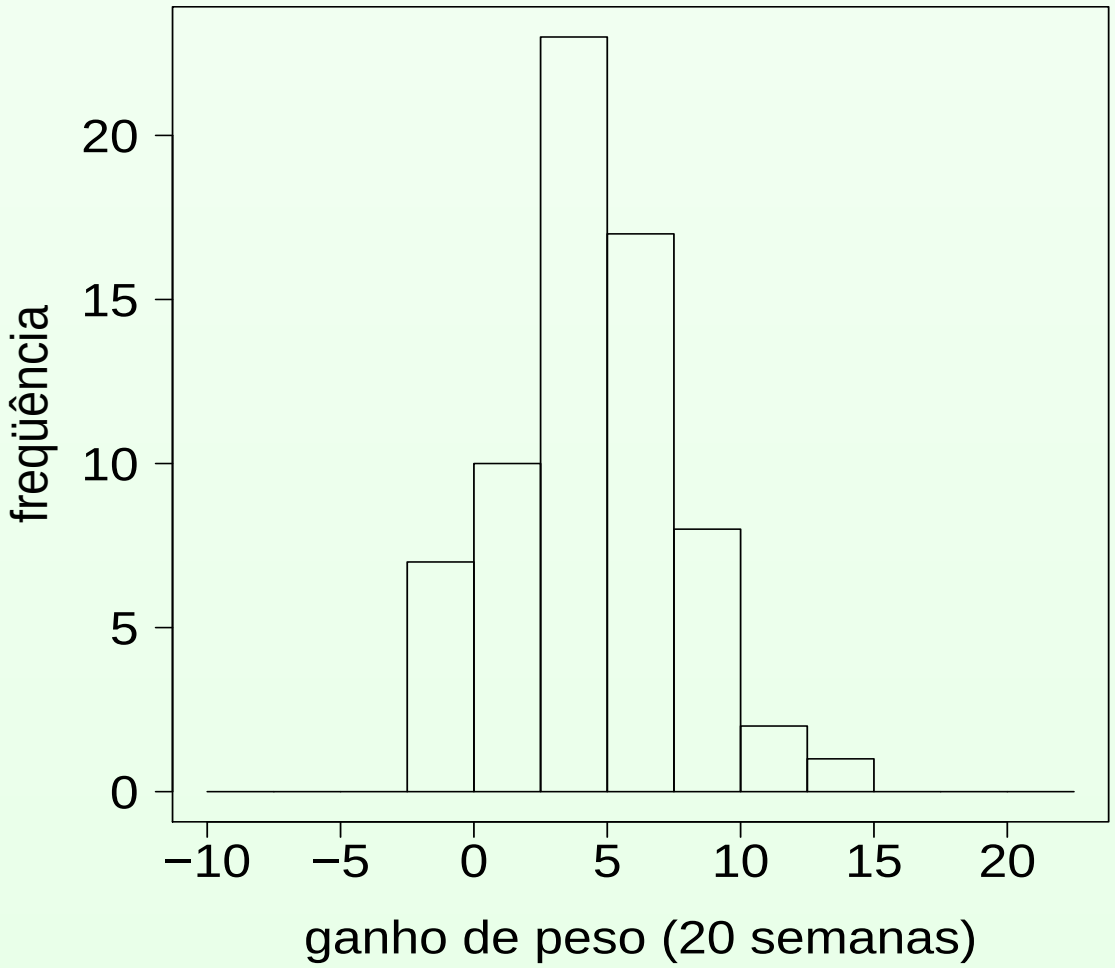
- Estudar o comportamento do ganho de peso ao longo da gestação, para avaliar a adequabilidade do instrumento;
- Verificar se existem diferenças no comportamento do ganho de peso, de acordo com o estado nutricional na concepção.

Dados sobre Ganho de Peso em Gestantes

Ganho de Peso x Tempo de Gestação (semanas)

Ganho de Peso (acumulado): Observado - Pré-Gravídico

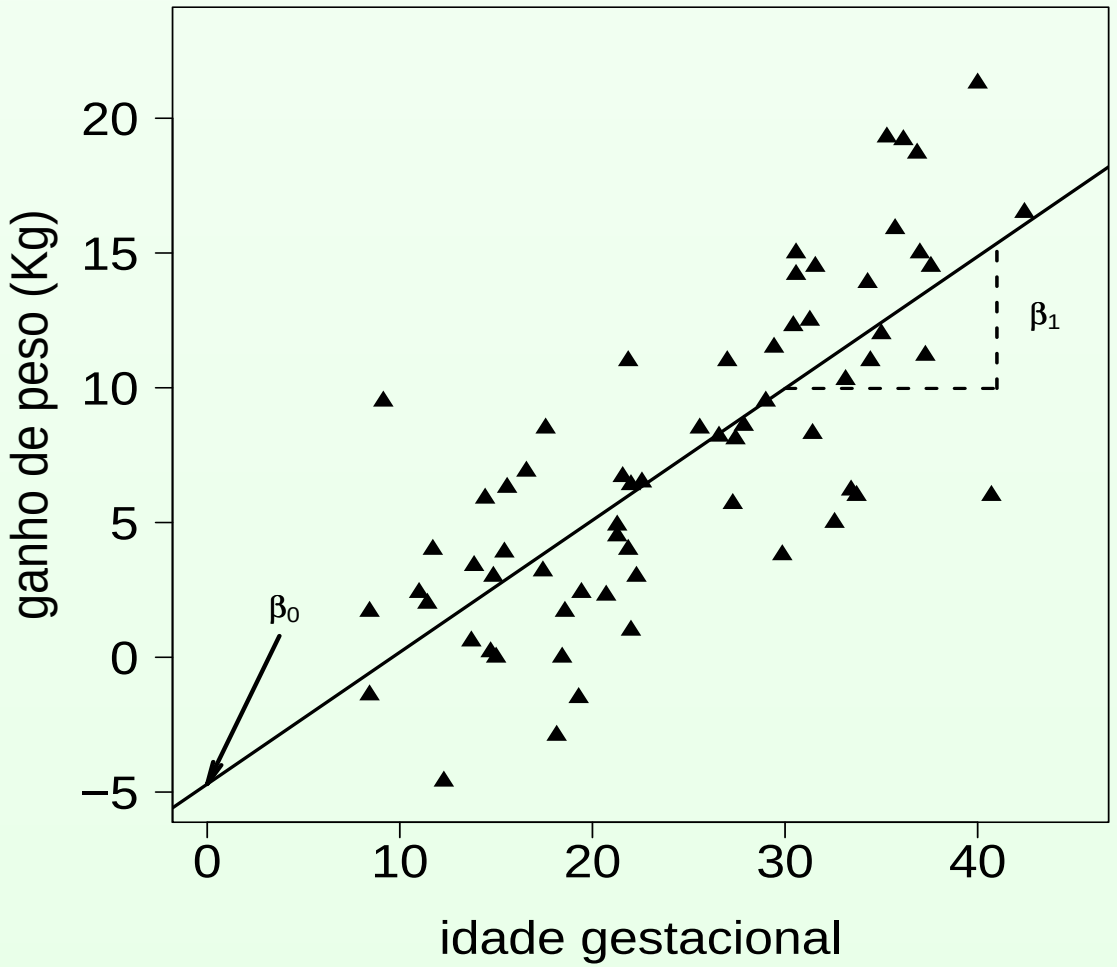
gest 1	3.3, 6.1, 8.5, 7.9, 11.8, 14.8
gest 1	8.571, 13.000, 17.571, 20.857, 24.571, 35.429
gest 2	2.0, 4.3, 8.0, 11.0, 12.2, 15.6
gest 2	14.429, 18.714, 23.000, 27.000, 31.000, 37.429
gest 3	4.5, 6.8, 8.1, 13.5, 15.0
gest 3	13.429, 18.857, 23.143, 34.143, 37.000
gest 4	5.5, 5.5, 8.0, 13.2, 13.9, 15.1
gest 4	12.571, 17.000, 21.429, 32.429, 34.286, 38.571
gest 5	2.0, 9.5, 12.2, 18.2, 19.3, 22.0
gest 5	13.714, 23.571, 27.714, 32.429, 35.286, 39.571
gest 6	1.8, 2.0, 4.5, 7.1, 9.5, 12.6, 13.5 gest 6



$$y \sim N(\mu, \sigma^2)$$

Histograma para o ganho de peso em gestantes.

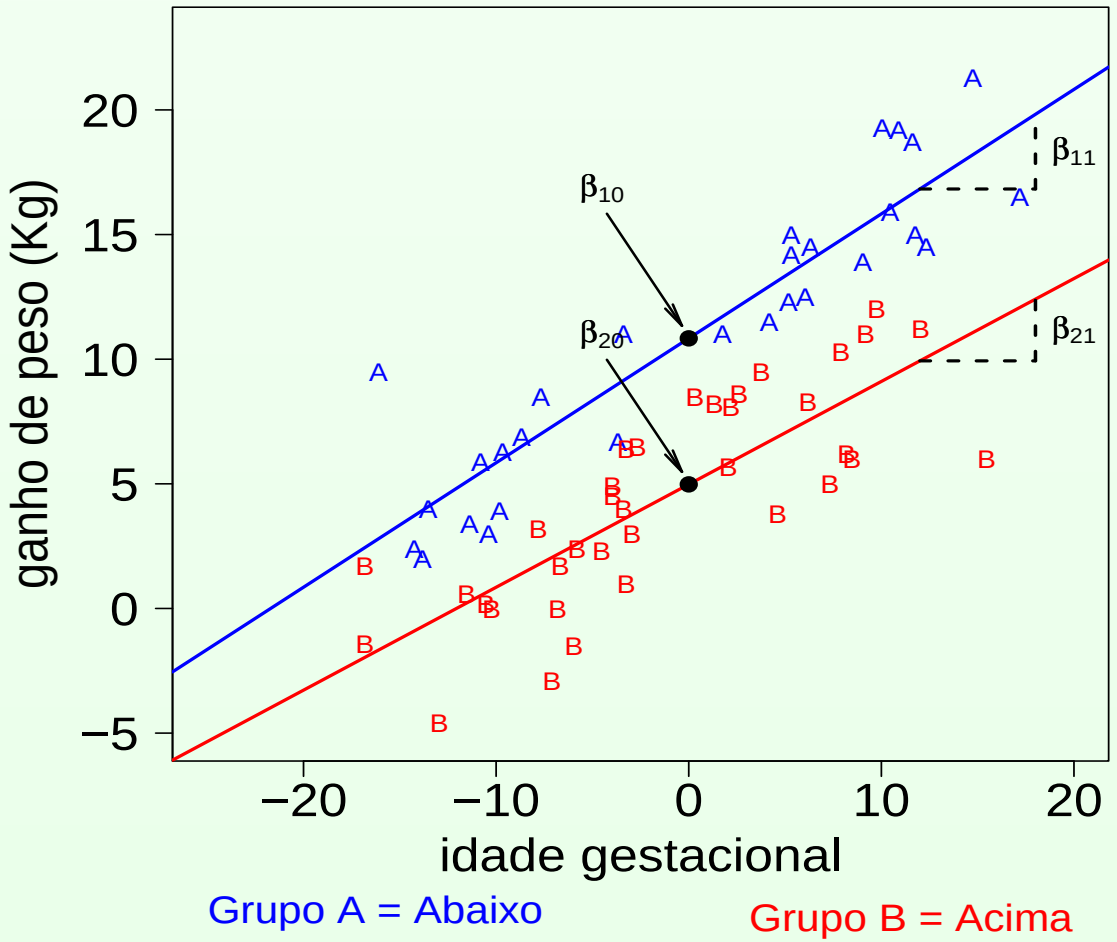
Análise do Ganho de Peso em Gestantes
Motivando Regressão Hierárquica
Modelos para a Análise do Ganho de Peso em Gestantes
Modelo hierárquico de dois estágios
Modelo hierárquico de três estágios
Modelo de Crescimento Linear



$$y_i \sim N(\mu_i, \sigma^2); \qquad \mu_i = \beta_0 + \beta_1 x_i$$

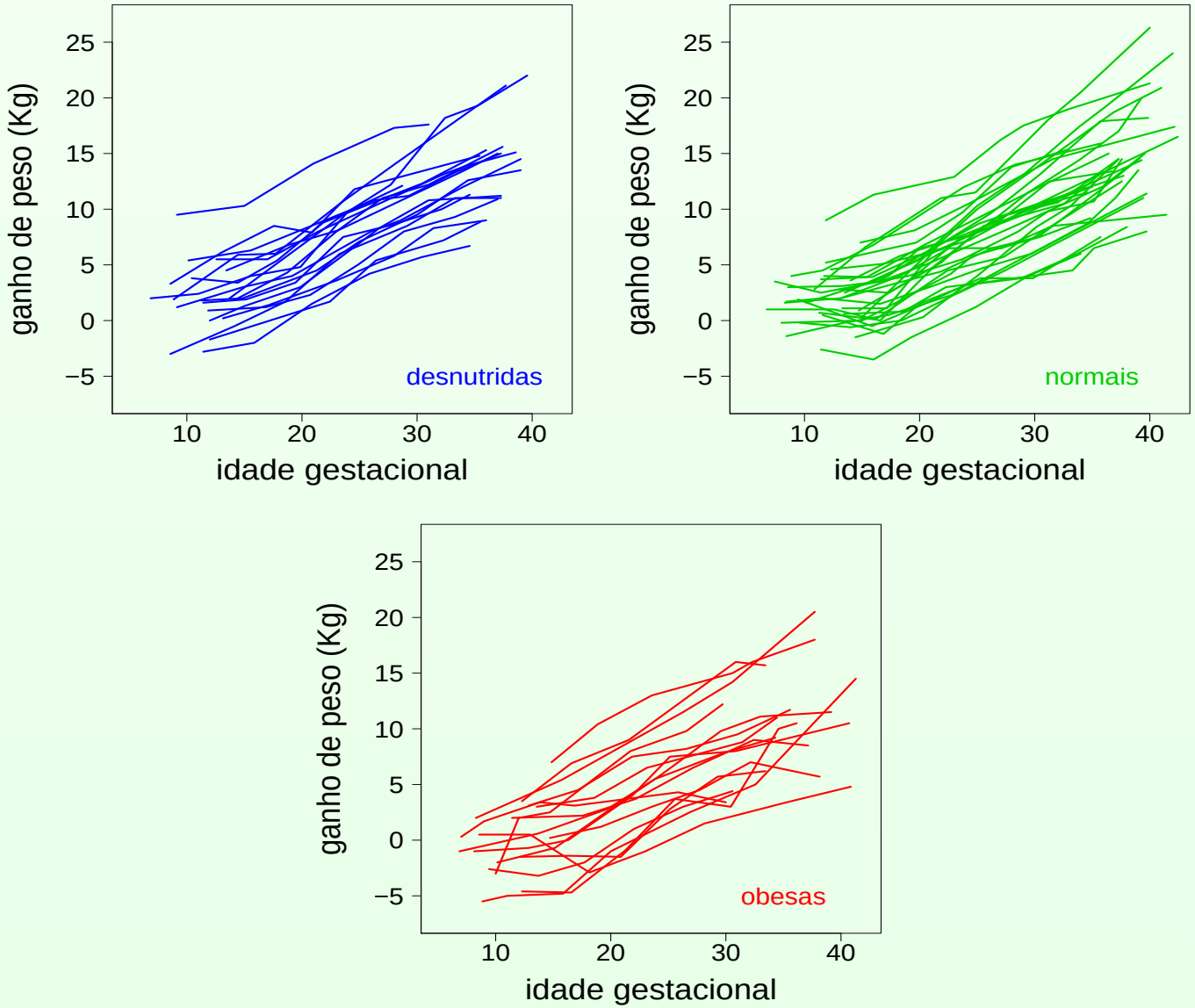
Relação entre o ganho de peso em gestantes e idade gestacional.

Análise do Ganho de Peso em Gestantes
Motivando Regressão Hierárquica
Modelos para a Análise do Ganho de Peso em Gestantes
Modelo hierárquico de dois estágios
Modelo hierárquico de três estágios
Modelo de Crescimento Linear



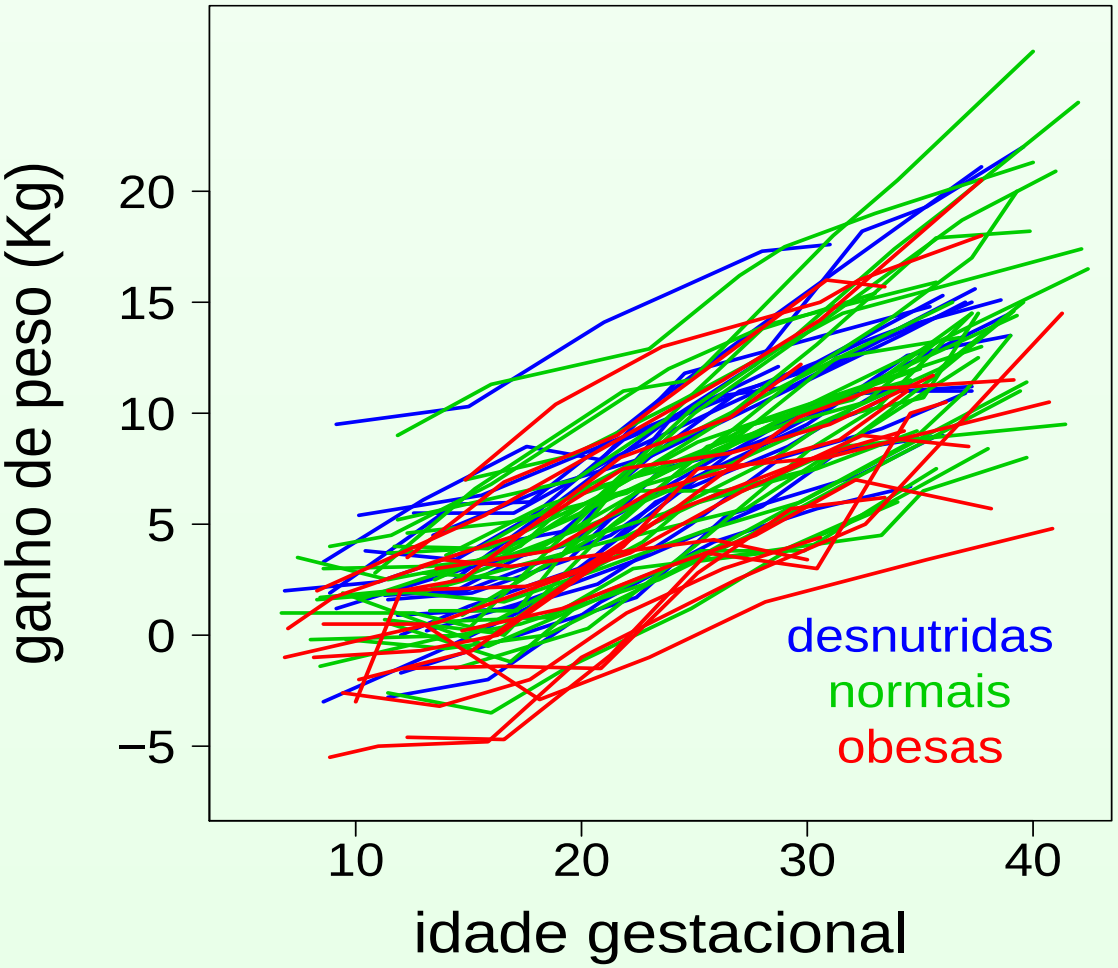
$$y_{gi} \sim N(\mu_{gi}, \sigma^2); \qquad \mu_{gi} = \beta_{0g} + \beta_{1g}(x_{gi} - \bar{x}_g)$$

Relação entre o ganho de peso e idade gestacional de acordo com estado nutricional na concepção.



Ganho de peso acumulado em função da idade gestacional para gestantes participantes de um programa pré-natal.

Análise do Ganho de Peso em Gestantes
Motivando Regressão Hierárquica
Modelos para a Análise do Ganho de Peso em Gestantes
Modelo hierárquico de dois estágios
Modelo hierárquico de três estágios
Modelo de Crescimento Linear



Ganho de peso acumulado em função da idade gestacional para gestantes participantes de um programa pré-natal.

Modelos para Análise do Ganho de Peso

Análise do Ganho de Peso em Gestantes
Motivando Regressão Hierárquica
Modelos para a Análise do Ganho de Peso em Gestantes
Modelo hierárquico de dois estágios
Modelo hierárquico de três estágios
Modelo de Crescimento Linear

Análise do Ganho de Peso em Gestantes

Análise do Ganho de Peso em Gestantes
Motivando Regressão Hierárquica
Modelos para a Análise do Ganho de Peso em Gestantes
Modelo hierárquico de dois estágios
Modelo hierárquico de três estágios
Modelo de Crescimento Linear

- Especificação do Modelo
  - Suposição: parâmetros descrevendo crescimentos individuais são permutáveis em torno de um parâmetro populacional comum.

- O modelo para a observação y_{it} no tempo x_{it}, $i = 1, 2, ..., n$, e $t = 1, 2..., t_i$ é dado por

$$\begin{aligned} y_{it} &\sim N(\mu_{it}, \sigma^2) \\ \mu_{it} &= \alpha_i + \beta_i x_{it}; \\ \begin{pmatrix} \alpha_i \\ \beta_i \end{pmatrix} &\sim N\left[\begin{pmatrix} \alpha \\ \beta \end{pmatrix}, \begin{pmatrix} \sigma^2_\alpha & 0 \\ 0 & \sigma^2_\beta \end{pmatrix}\right], \end{aligned} \tag{1}$$

- y_{it} ganho de peso até o tempo x_{it} para a i-ésima gestante;
- t_i número de informações para a i-ésima gestante.

Análise do Ganho de Peso em Gestantes

- Especificação do Modelo
  - Suposição: parâmetros descrevendo crescimentos individuais são permutáveis em torno de um parâmetro populacional comum.

- O modelo para a observação y_{it} no tempo x_{it}, $i = 1, 2, ..., n$, e $t = 1, 2..., t_i$ é dado por

$$\begin{aligned} y_{it} &\sim N(\mu_{it}, \sigma^2) \\ \mu_{it} &= \alpha_i + \beta_i x_{it}; \\ \begin{pmatrix} \alpha_i \\ \beta_i \end{pmatrix} &\sim N\left[\begin{pmatrix} \alpha \\ \beta \end{pmatrix}, \begin{pmatrix} \sigma^2_\alpha & 0 \\ 0 & \sigma^2_\beta \end{pmatrix}\right], \end{aligned} \tag{1}$$

- y_{it} ganho de peso até o tempo x_{it} para a i-ésima gestante;
- t_i número de informações para a i-ésima gestante.

Análise do Ganho de Peso em Gestantes
Motivando Regressão Hierárquica
Modelos para a Análise do Ganho de Peso em Gestantes
Modelo hierárquico de dois estágios
Modelo hierárquico de três estágios
Modelo de Crescimento Linear

Análise do Ganho de Peso em Gestantes

Análise do Ganho de Peso em Gestantes
Motivando Regressão Hierárquica
Modelos para a Análise do Ganho de Peso em Gestantes
Modelo hierárquico de dois estágios
Modelo hierárquico de três estágios
Modelo de Crescimento Linear

- Especificação do Modelo
  - Suposição: parâmetros descrevendo crescimentos individuais são permutáveis em torno de um parâmetro populacional comum.

- O modelo para a observação y_{it} no tempo x_{it}, $i = 1, 2, ..., n$, e $t = 1, 2..., t_i$ é dado por

$$\begin{aligned} y_{it} &\sim N(\mu_{it}, \sigma^2) \\ \mu_{it} &= \alpha_i + \beta_i x_{it}; \\ \begin{pmatrix} \alpha_i \\ \beta_i \end{pmatrix} &\sim N\left[\begin{pmatrix} \alpha \\ \beta \end{pmatrix}, \begin{pmatrix} \sigma^2_\alpha & 0 \\ 0 & \sigma^2_\beta \end{pmatrix}\right], \end{aligned} \tag{1}$$

- y_{it} ganho de peso até o tempo x_{it} para a i-ésima gestante;
- t_i número de informações para a i-ésima gestante.

Análise do Ganho de Peso em Gestantes

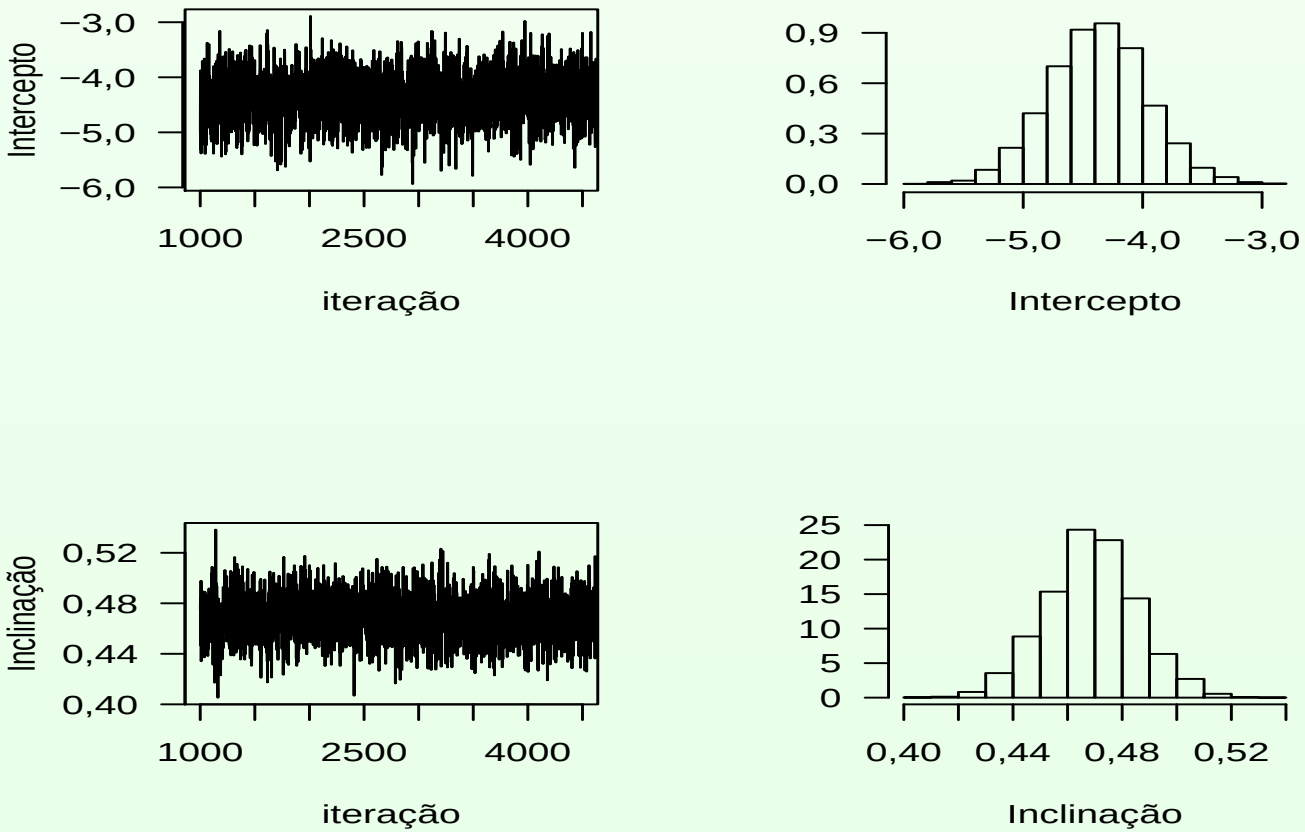


Figura 3.16: Trajetórias e histogramas gerados via amostrador de Gibbs para as distribuições a posteriori de α e β^2 sob o modelo em (1).

Análise do Ganho de Peso em Gestantes
Motivando Regressão Hierárquica
Modelos para a Análise do Ganho de Peso em Gestantes
Modelo hierárquico de dois estágios
Modelo hierárquico de três estágios
Modelo de Crescimento Linear

Análise do Ganho de Peso em Gestantes

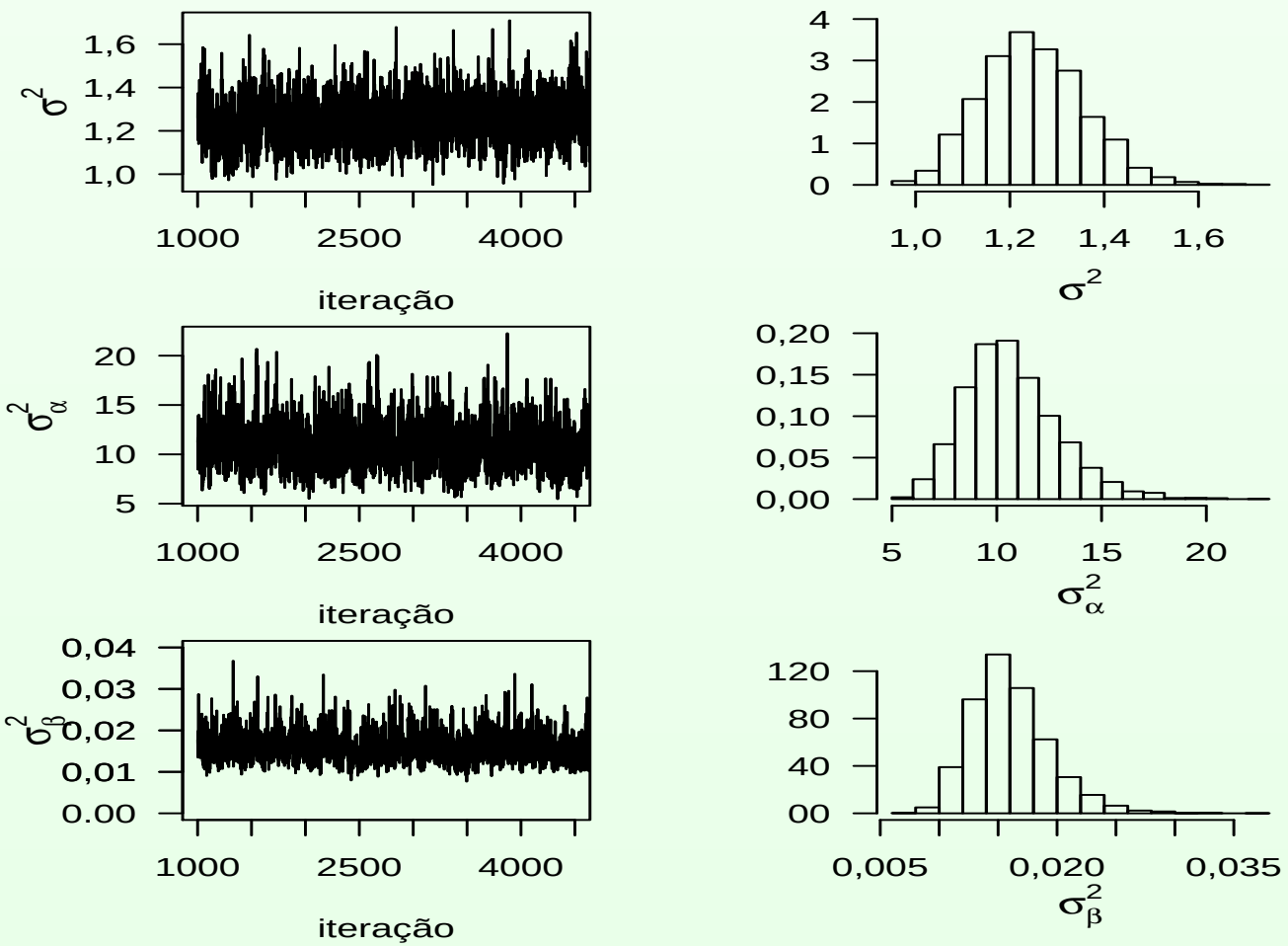


Figura 3.17: Trajetórias e histogramas gerados via amostrador de Gibbs para as distribuições a posteriori de σ^2, σ^2_α, σ^2_β sob o modelo (1).

Análise do Ganho de Peso em Gestantes
Motivando Regressão Hierárquica
Modelos para a Análise do Ganho de Peso em Gestantes
Modelo hierárquico de dois estágios
Modelo hierárquico de três estágios
Modelo de Crescimento Linear

Análise do Ganho de Peso em Gestantes

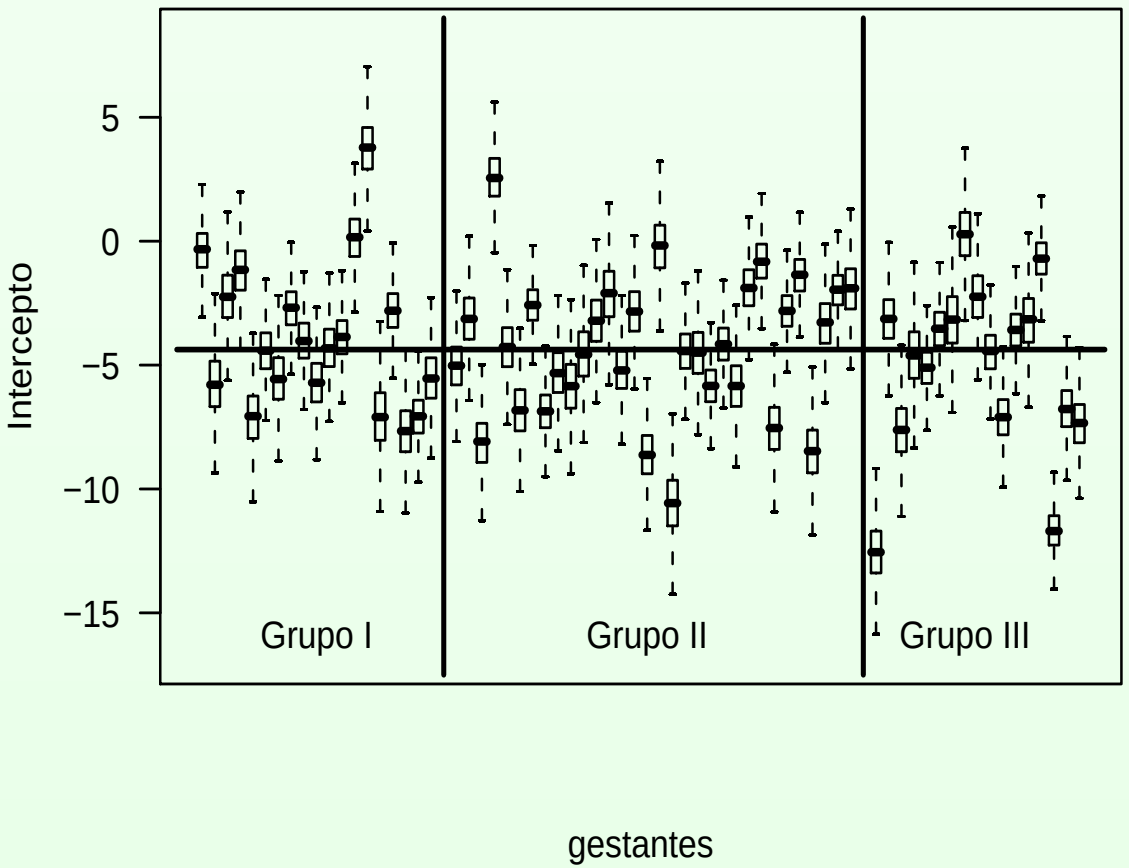


Figura 3.18: Boxplots para as Distribuições a Posteriori dos Parâmetros α_i (Intercepto) por Gestante, no modelo em (1) (linha horizontal - média a posteriori de α)

Análise do Ganho de Peso em Gestantes
Motivando Regressão Hierárquica
Modelos para a Análise do Ganho de Peso em Gestantes
Modelo hierárquico de dois estágios
Modelo hierárquico de três estágios
Modelo de Crescimento Linear

Análise do Ganho de Peso em Gestantes

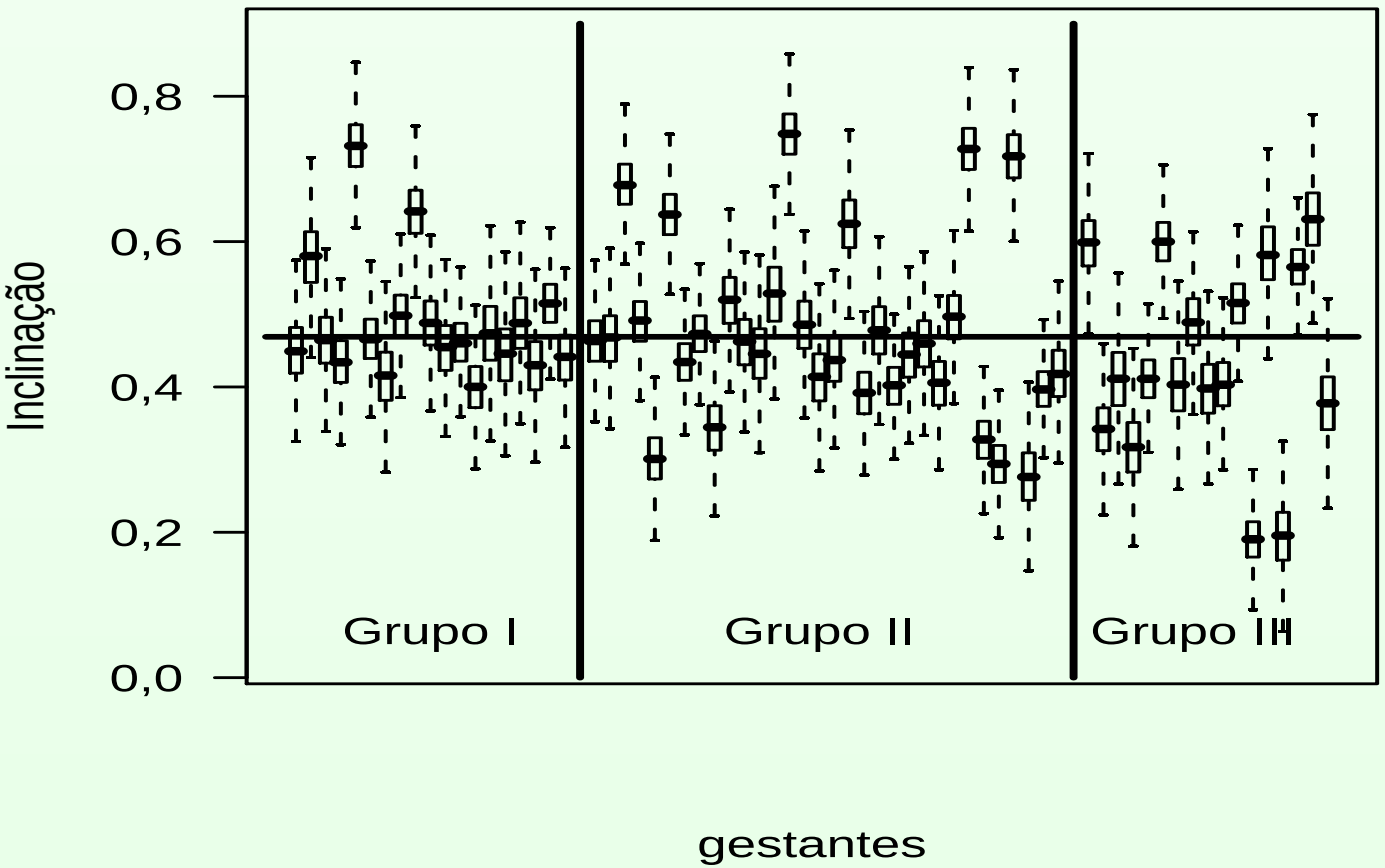


Figura 3.19: Boxplots para as Distribuições a Posteriori dos Parâmetros β_i (Intercepto) por Gestante, no modelo em (1) (linha horizontal - média a posteriori de β)

Análise do Ganho de Peso em Gestantes
Motivando Regressão Hierárquica
Modelos para a Análise do Ganho de Peso em Gestantes
Modelo hierárquico de dois estágios
Modelo hierárquico de três estágios
Modelo de Crescimento Linear

Modelo hierárquico de três estágios

- Modelo (1): gestantes consideradas individualmente;
- Objetivo: estudar o comportamento do ganho de peso conforme estado nutricional na concepção;
- Modelo para a observação y_{git}, no tempo x_{git}, para a i-ésima gestante do g-ésimo grupo é, especificado por

$$\begin{aligned} y_{git} &\sim N(\mu_{git}, \sigma_g^2); \\ \mu_{git} &= \alpha_{gi} + \beta_{gi} x_{git}; \\ \begin{pmatrix} \alpha_{gi} \\ \beta_{gi} \end{pmatrix} &\sim N\left[\begin{pmatrix} \alpha_g \\ \beta_g \end{pmatrix}, \begin{pmatrix} \sigma^2_{\alpha g} & 0 \\ 0 & \sigma^2_{\beta g} \end{pmatrix} \right], \quad (2) \\ \begin{pmatrix} \alpha_g \\ \beta_g \end{pmatrix} &\sim N\left[\begin{pmatrix} \alpha \\ \beta \end{pmatrix}, \begin{pmatrix} \sigma^2_{\alpha} & 0 \\ 0 & \sigma^2_{\beta} \end{pmatrix} \right], \end{aligned}$$

onde $i = 1, 2, ..., n_g$, $t = 1, 2..., t_{gi}$, $g = 1, ..., G$ e $n = n_1 + n_2 + n_3$.

Modelo hierárquico de três estágios

- Modelo (1): gestantes consideradas individualmente;
- Objetivo: estudar o comportamento do ganho de peso conforme estado nutricional na concepção;
- Modelo para a observação y_{git}, no tempo x_{git}, para a i-ésima gestante do g-ésimo grupo é, especificado por

$$
\begin{aligned}
y_{git} &\sim N(\mu_{git}, \sigma_g^2); \\
\mu_{git} &= \alpha_{gi} + \beta_{gi} x_{git}; \\
\begin{pmatrix} \alpha_{gi} \\ \beta_{gi} \end{pmatrix} &\sim N\left[\begin{pmatrix} \alpha_g \\ \beta_g \end{pmatrix}, \begin{pmatrix} \sigma^2_{\alpha g} & 0 \\ 0 & \sigma^2_{\beta g} \end{pmatrix} \right], \quad (2) \\
\begin{pmatrix} \alpha_g \\ \beta_g \end{pmatrix} &\sim N\left[\begin{pmatrix} \alpha \\ \beta \end{pmatrix}, \begin{pmatrix} \sigma^2_{\alpha} & 0 \\ 0 & \sigma^2_{\beta} \end{pmatrix} \right],
\end{aligned}
$$

onde $i = 1, 2, ..., n_g$, $t = 1, 2..., t_{gi}$, $g = 1, ..., G$ e $n = n_1 + n_2 + n_3$.

Análise do Ganho de Peso em Gestantes
Motivando Regressão Hierárquica
Modelos para a Análise do Ganho de Peso em Gestantes
Modelo hierárquico de dois estágios
Modelo hierárquico de três estágios
Modelo de Crescimento Linear

Modelo hierárquico de três estágios

- Modelo (1): gestantes consideradas individualmente;
- Objetivo: estudar o comportamento do ganho de peso conforme estado nutricional na concepção;
- Modelo para a observação y_{git}, no tempo x_{git}, para a i-ésima gestante do g-ésimo grupo é, especificado por

$$\begin{aligned} y_{git} &\sim N(\mu_{git}, \sigma_g^2); \\ \mu_{git} &= \alpha_{gi} + \beta_{gi} x_{git}; \\ \begin{pmatrix} \alpha_{gi} \\ \beta_{gi} \end{pmatrix} &\sim N\left[\begin{pmatrix} \alpha_g \\ \beta_g \end{pmatrix}, \begin{pmatrix} \sigma^2_{\alpha g} & 0 \\ 0 & \sigma^2_{\beta g} \end{pmatrix} \right], \qquad (2) \\ \begin{pmatrix} \alpha_g \\ \beta_g \end{pmatrix} &\sim N\left[\begin{pmatrix} \alpha \\ \beta \end{pmatrix}, \begin{pmatrix} \sigma^2_{\alpha} & 0 \\ 0 & \sigma^2_{\beta} \end{pmatrix} \right], \end{aligned}$$

onde $i = 1, 2, ..., n_g$, $t = 1, 2..., t_{gi}$, $g = 1, ..., G$ e $n = n_1 + n_2 + n_3$.

Análise do Ganho de Peso em Gestantes
Motivando Regressão Hierárquica
Modelos para a Análise do Ganho de Peso em Gestantes
Modelo hierárquico de dois estágios
Modelo hierárquico de três estágios
Modelo de Crescimento Linear

Análise do Ganho de Peso em Gestantes
Motivando Regressão Hierárquica
Modelos para a Análise do Ganho de Peso em Gestantes
Modelo hierárquico de dois estágios
Modelo hierárquico de três estágios
Modelo de Crescimento Linear

Tabela 3.17: Ganhos Acumulados de Peso Estimados através do modelo em (2).

Semanas	Grupo I	Grupo II	Grupo III	Populacional
10	0,763	0,485	-0,069	0,486
15	3,201	2,885	2,110	2,804
20	5,639	5,524	4,289	5,122
25	8,077	7,594	6,468	7,440
30	10,515	9,963	8,647	9.758
35	12,953	12,333	10,826	12,076

Análise do Ganho de Peso em Gestantes

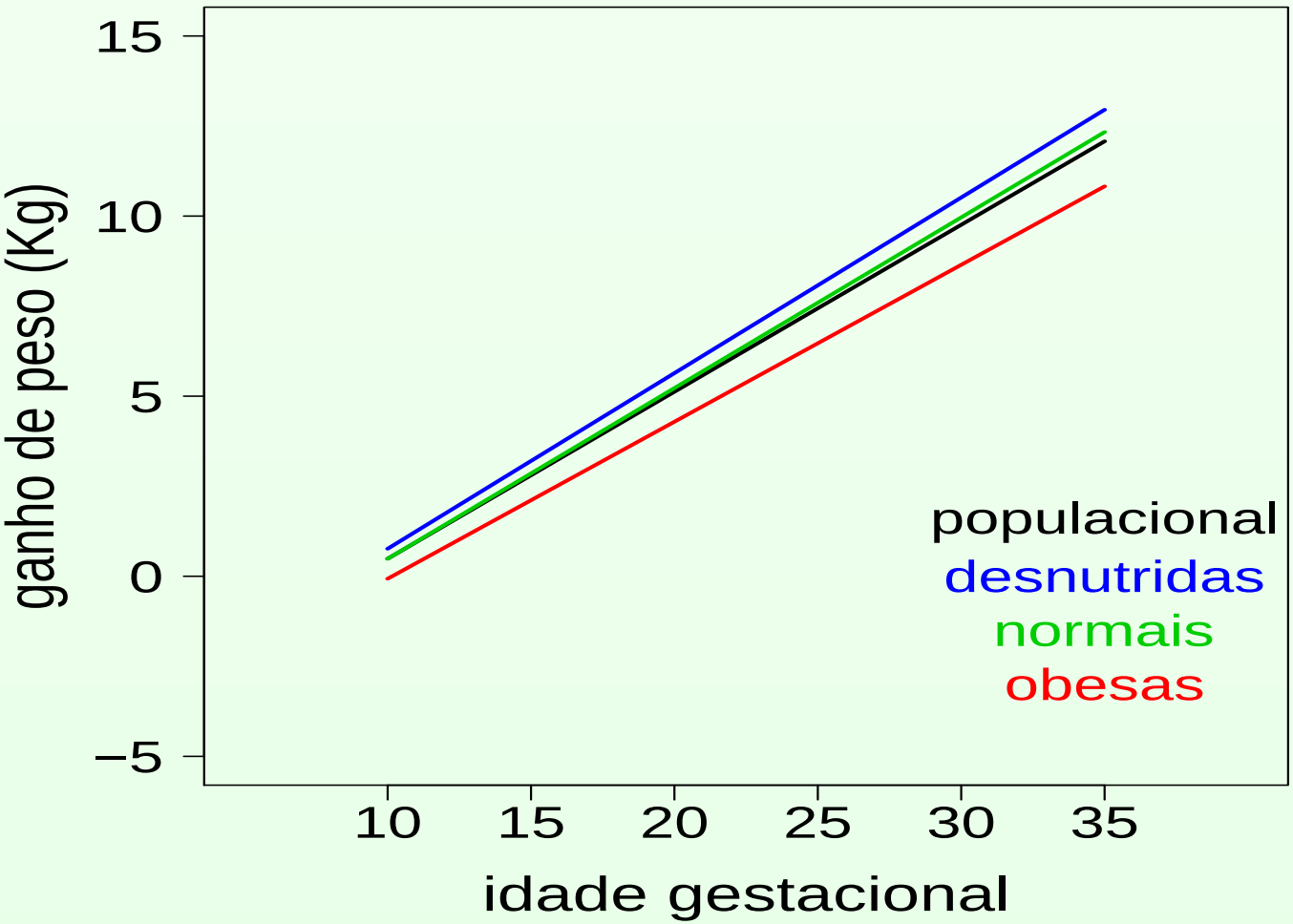


Figura Tab. 3.17: Ganhos Acumulados de Peso Estimados através do modelo em (2).

Análise do Ganho de Peso em Gestantes
Motivando Regressão Hierárquica
Modelos para a Análise do Ganho de Peso em Gestantes
Modelo hierárquico de dois estágios
Modelo hierárquico de três estágios
Modelo de Crescimento Linear

Análise do Ganho de Peso em Gestantes

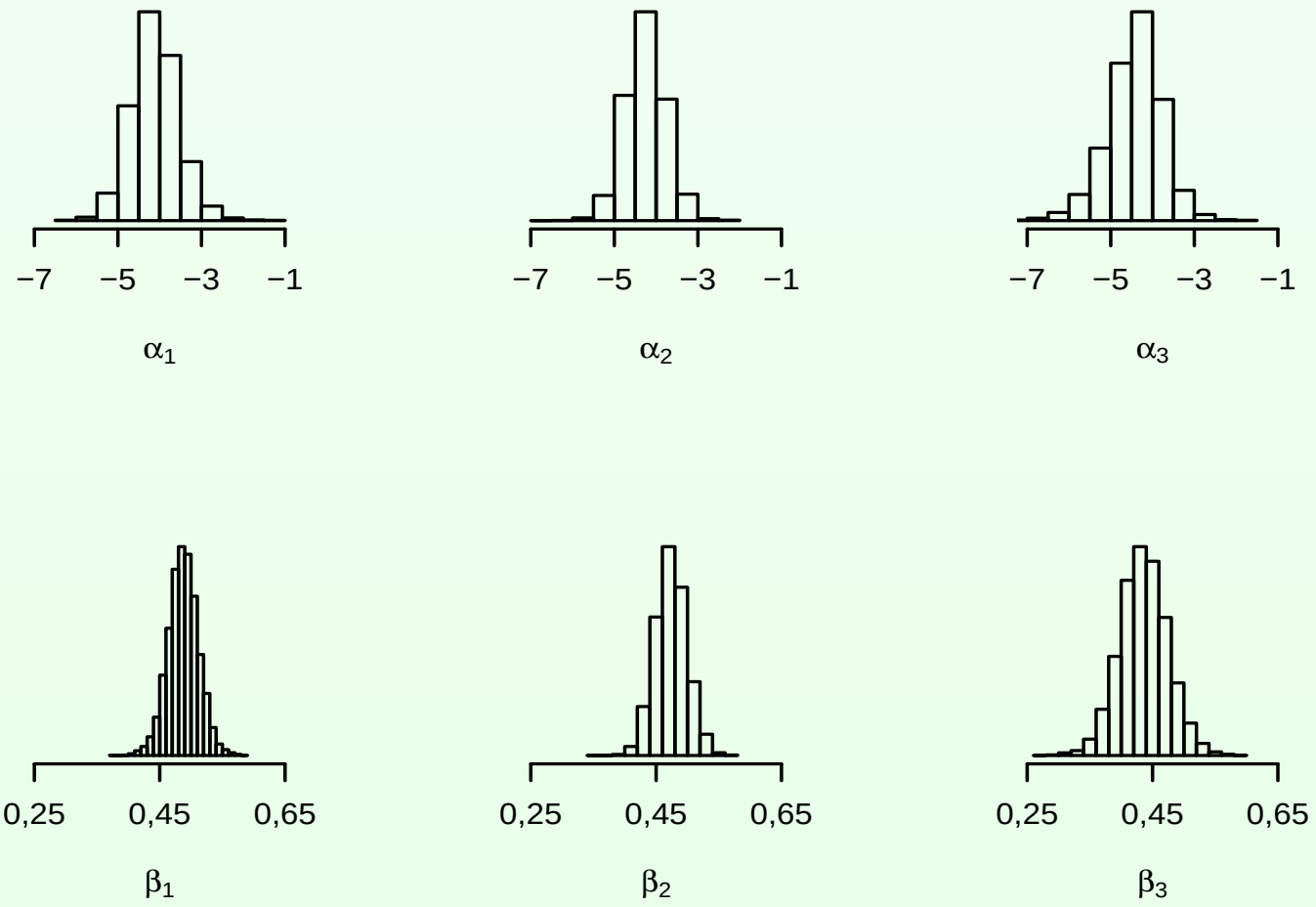


Figura 3.20: Histrogramas das distribuições a Posteriori para os Parâmetros: $\alpha_g, g = 1, 2, 3$ e $\beta_g, g = 1, 2, 3$ sob o modelo em (2).

Análise do Ganho de Peso em Gestantes
Motivando Regressão Hierárquica
Modelos para a Análise do Ganho de Peso em Gestantes
Modelo hierárquico de dois estágios
Modelo hierárquico de três estágios
Modelo de Crescimento Linear

Análise do Ganho de Peso em Gestantes

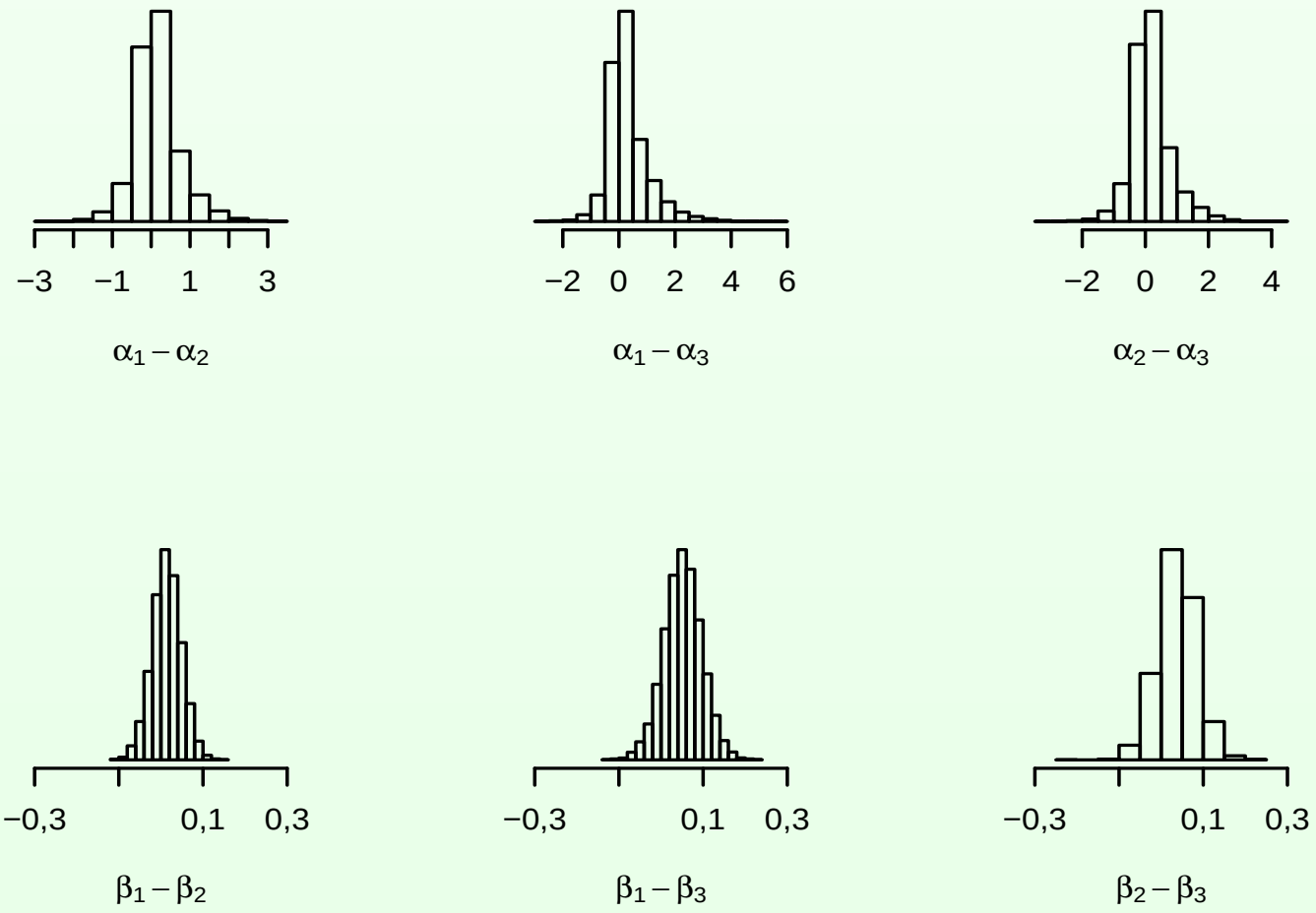
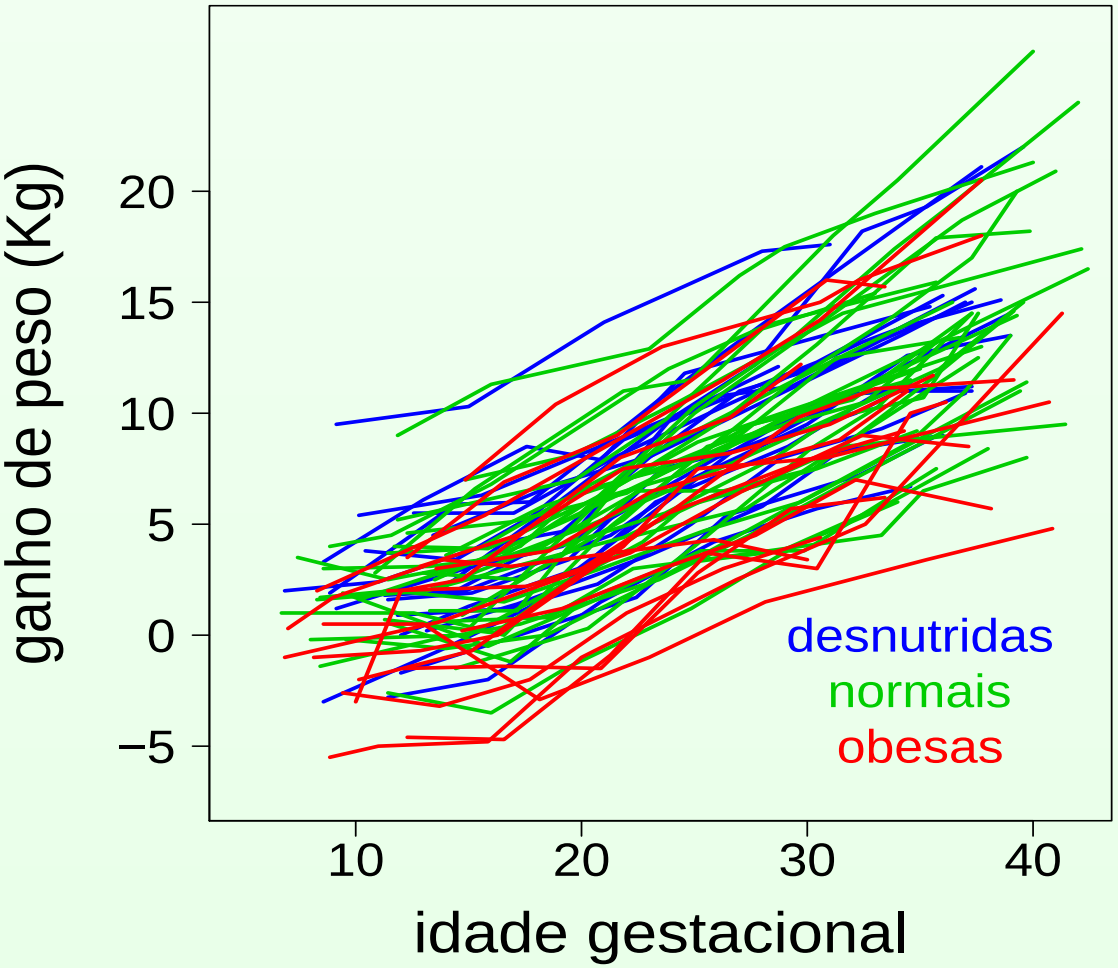


Figura 3.21: Histrogramas das distribuições a Posteriori para as Diferenças entre os Parâmetros: $\alpha_g, g = 1, 2, 3$ e $\beta_g, g = 1, 2, 3$ sob o modelo em (2)

Análise do Ganho de Peso em Gestantes
Motivando Regressão Hierárquica
Modelos para a Análise do Ganho de Peso em Gestantes
Modelo hierárquico de dois estágios
Modelo hierárquico de três estágios
Modelo de Crescimento Linear



Ganho de peso acumulado em função da idade gestacional para gestantes participantes de um programa pré-natal.

Análise do Ganho de Peso em Gestantes
Motivando Regressão Hierárquica
Modelos para a Análise do Ganho de Peso em Gestantes
Modelo hierárquico de dois estágios
Modelo hierárquico de três estágios
Modelo de Crescimento Linear

Modelos Lineares Dinâmicos Hierárquicos

- Versão dinâmica para o modelo hierárquico de dois estágios em (1), onde a tendência global é $\mu_t = \alpha + \beta x_t$.

$$\begin{aligned} y_{it} &\sim N(\mu_{it}, \sigma^2), \\ \mu_{it} &\sim N(\mu_t, \sigma^2_\mu), \end{aligned} \tag{3}$$

com a evolução para μ_t como (4).

$$\begin{aligned} \mu_t &= \mu_{t-1} + \gamma_t + w_{1t}, \\ \gamma_t &= \gamma_{t-1} + w_{2t} \end{aligned} \tag{4}$$

com $w_{jt} \sim N(0, W_j)$, $j = 1, 2$.

- Dados interpolados: 10, 15, 20, 25, 30 e 35 semanas.
- Problema de identificabilidade: $\sigma^2 = \rho^2 \sigma^2_\mu$, $\rho^2 \sim Ga(2; 0,5)$.

Modelos Lineares Dinâmicos Hierárquicos

- Versão dinâmica para o modelo hierárquico de dois estágios em (1), onde a tendência global é $\mu_t = \alpha + \beta x_t$.

$$\begin{aligned} y_{it} &\sim N(\mu_{it}, \sigma^2), \\ \mu_{it} &\sim N(\mu_t, \sigma^2_\mu), \end{aligned} \tag{3}$$

com a evolução para μ_t como (4).

$$\begin{aligned} \mu_t &= \mu_{t-1} + \gamma_t + w_{1t}, \\ \gamma_t &= \gamma_{t-1} + w_{2t} \end{aligned} \tag{4}$$

com $w_{jt} \sim N(0, W_j)$, $j = 1, 2$.

- Dados interpolados: 10, 15, 20, 25, 30 e 35 semanas.
- Problema de identificabilidade: $\sigma^2 = \rho^2 \sigma^2_\mu$, $\rho^2 \sim Ga(2; 0,5)$.

Modelos Lineares Dinâmicos Hierárquicos

Análise do Ganho de Peso em Gestantes
Motivando Regressão Hierárquica
Modelos para a Análise do Ganho de Peso em Gestantes
Modelo hierárquico de dois estágios
Modelo hierárquico de três estágios
Modelo de Crescimento Linear

- Versão dinâmica para o modelo hierárquico de dois estágios em (1), onde a tendência global é $\mu_t = \alpha + \beta x_t$.

$$\begin{aligned} y_{it} &\sim N(\mu_{it}, \sigma^2), \\ \mu_{it} &\sim N(\mu_t, \sigma^2_\mu), \end{aligned} \tag{3}$$

com a evolução para μ_t como (4).

$$\begin{aligned} \mu_t &= \mu_{t-1} + \gamma_t + w_{1t}, \\ \gamma_t &= \gamma_{t-1} + w_{2t} \end{aligned} \tag{4}$$

com $w_{jt} \sim N(0, W_j)$, $j = 1, 2$.

- Dados interpolados: 10, 15, 20, 25, 30 e 35 semanas.
- Problema de identificabilidade: $\sigma^2 = \rho^2 \sigma^2_\mu$, $\rho^2 \sim Ga(2; 0,5)$.

Tabela 5.2: Média a Posteriori para o Ganho de Peso Acumulado (μ_t) e Taxa de Crescimento (γ_t) em (4).

Semanas	Ganho de Peso (d.p.)	Taxa
10	1,069 (0,359)	0,784 (0,513)
15	2,411 (0,366)	1,341 (0,525)
20	4,514 (0,362)	1,943 (0,543)
25	7,233 (0,362)	2,382 (0,556)
30	9,734 (0,363)	2,397 (0,563)
35	12,030 (0,410)	2,330 (0,651)

Tabela 5.3: Valores Estimados para Parâmetros de Variância.

Parâmetro	Média (d.p.)	I.C. 95 %
σ^2	8,540 (2,226)	(3,208; 11,910)
σ^2_μ	3,838 (2,156)	(1,103; 9,185)
ρ^2	3,390 (2,602)	(0,354; 10,030)
W_1	0,537 (0,437)	(0,156; 1,586)
W_2	0,726 (0,518)	(0,221; 2,061)

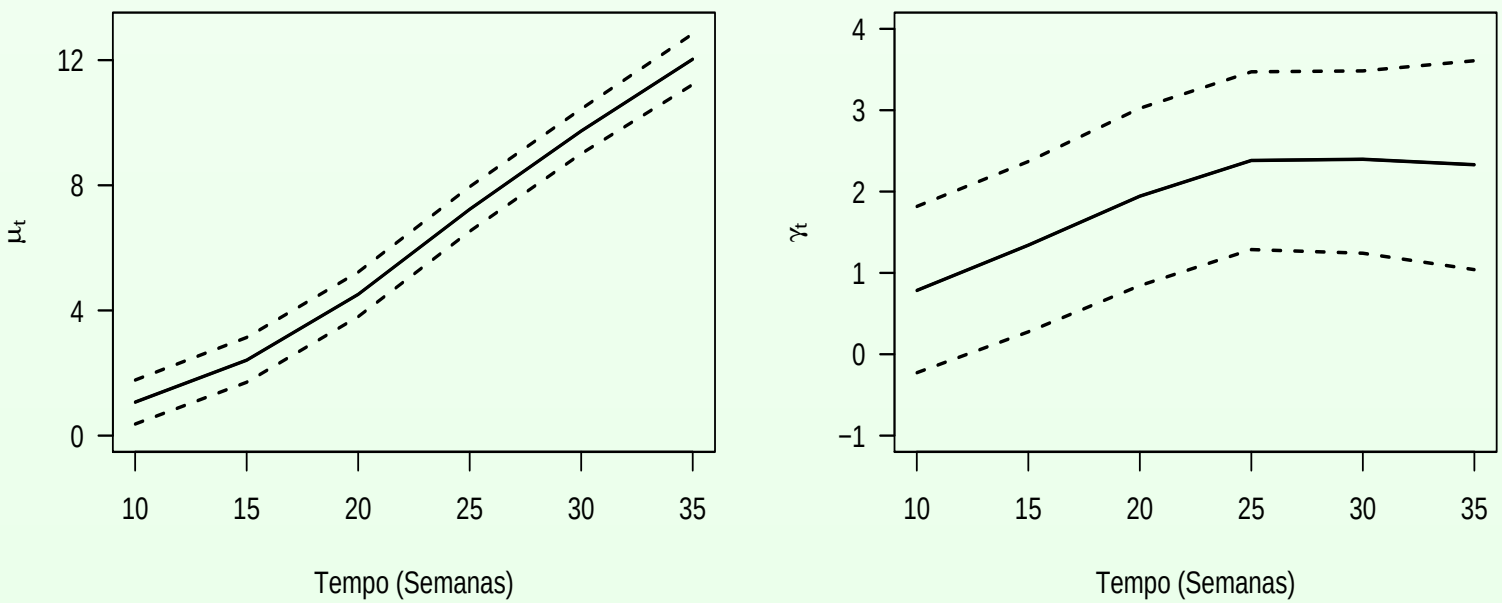
Análise do Ganho de Peso em Gestantes
Motivando Regressão Hierárquica
Modelos para a Análise do Ganho de Peso em Gestantes
Modelo hierárquico de dois estágios
Modelo hierárquico de três estágios
Modelo de Crescimento Linear

Tabela 5.2: Média a Posteriori para o Ganho de Peso Acumulado (μ_t) e Taxa de Crescimento (γ_t) em (4).

Semanas	Ganho de Peso (d.p.)	Taxa
10	1,069 (0,359)	0,784 (0,513)
15	2,411 (0,366)	1,341 (0,525)
20	4,514 (0,362)	1,943 (0,543)
25	7,233 (0,362)	2,382 (0,556)
30	9,734 (0,363)	2,397 (0,563)
35	12,030 (0,410)	2,330 (0,651)

Tabela 5.3: Valores Estimados para Parâmetros de Variância.

Parâmetro	Média (d.p.)	I.C. 95 %
σ^2	8,540 (2,226)	(3,208; 11,910)
σ^2_μ	3,838 (2,156)	(1,103; 9,185)
ρ^2	3,390 (2,602)	(0,354; 10,030)
W_1	0,537 (0,437)	(0,156; 1,586)
W_2	0,726 (0,518)	(0,221; 2,061)



(a) (b)

Figura 5.7: Média a Posteriori para: (a) Ganho Médio de Peso Acumulado; (b) Taxa de Crescimento com $10, 15, \cdots, 35$ Semanas de Gestação, para o modelo em (4); (linha cheia representa a média e tracejada os IC 95%).

- Resultados: duas cadeias de 120.000 iterações, descartadas as 20.000 iniciais e espaçamento de 20.

Modelos para Dados de Contagem

Modelos para Dados de Contagem
Modelo de Poisson
Aplicações
Relação entre o hábito de fumar e mortes por doenças coronarianas

Exemplo de Modelo para Dado de Contagem

- A distribuição de Poisson é frequentemente usada para modelar dados de contagem;
- Sua distribuição de probabilidade pode ser escrita como:

$$p(y|\mu) = \frac{\mu^y \exp^{-\mu}}{y!}, \qquad y = 0, 1, 2, \cdots$$

onde $\mu = E(Y)$ representa o número médio de ocorrências;

- O efeito de variáveis explicativas sobre Y pode ser modelado através do parâmetro μ;
- Considere $Y_1, Y_2, \cdots, Y_N$ v.as. i.i.d., com Y_i denotando o número de eventos obsevados em n_i expostos, de acordo com uma covariável padrão;

$$E(Y_i) = \mu_i = n_i\theta_i, \qquad i = 1, 2, \cdots, N.$$

Exemplo de Modelo para Dado de Contagem

- A distribuição de Poisson é frequentemente usada para modelar dados de contagem;
- Sua distribuição de probabilidade pode ser escrita como:

$$p(y|\mu) = \frac{\mu^y \exp^{-\mu}}{y!}, \qquad y = 0, 1, 2, \cdots$$

onde $\mu = E(Y)$ representa o número médio de ocorrências;

- O efeito de variáveis explicativas sobre Y pode ser modelado através do parâmetro μ;
- Considere $Y_1, Y_2, \cdots, Y_N$ v.as. i.i.d., com Y_i denotando o número de eventos obsevados em n_i expostos, de acordo com uma covariável padrão;

$$E(Y_i) = \mu_i = n_i\theta_i, \qquad i = 1, 2, \cdots, N.$$

Exemplo de Modelo para Dado de Contagem

- A distribuição de Poisson é frequentemente usada para modelar dados de contagem;
- Sua distribuição de probabilidade pode ser escrita como:

$$p(y|\mu) = \frac{\mu^y \exp^{-\mu}}{y!}, \qquad y = 0, 1, 2, \cdots$$

onde $\mu = E(Y)$ representa o número médio de ocorrências;

- O efeito de variáveis explicativas sobre Y pode ser modelado através do parâmetro μ;
- Considere $Y_1, Y_2, \cdots, Y_N$ v.as. i.i.d., com Y_i denotando o número de eventos obsevados em n_i expostos, de acordo com uma covariável padrão;

$$E(Y_i) = \mu_i = n_i\theta_i, \qquad i = 1, 2, \cdots, N.$$

Exemplo de Modelo para Dado de Contagem

- A distribuição de Poisson é frequentemente usada para modelar dados de contagem;
- Sua distribuição de probabilidade pode ser escrita como:

$$p(y|\mu) = \frac{\mu^y \exp^{-\mu}}{y!}, \qquad y = 0, 1, 2, \cdots$$

onde $\mu = E(Y)$ representa o número médio de ocorrências;

- O efeito de variáveis explicativas sobre Y pode ser modelado através do parâmetro μ;
- Considere $Y_1, Y_2, \cdots, Y_N$ v.as. i.i.d., com Y_i denotando o número de eventos obsevados em n_i expostos, de acordo com uma covariável padrão;

$$E(Y_i) = \mu_i = n_i\theta_i, \qquad i = 1, 2, \cdots, N.$$

Exemplo de Modelo para Dados de Contagem

- Por exemplo, suponha que Y_i seja a taxa de óbito por câncer.
- A dependência de θ_i em relação às variáveis explicativas é usualmente modelado por:

$$\theta_i = \exp(\mathbf{x}_i'\beta).$$

- Portanto, o modelo é definido por

$$Y_i \sim Poisson(\mu_i); \qquad E(Y_i) = \mu_i = n_i \exp(\mathbf{x}_i'\beta).$$

- Logo, a função de ligação natural é a função logaritmica

$$log\mu_i = logn_i + \mathbf{x}_i'\beta.$$

Aplicação: Dados da Literatura

- Aplicação (DOBSON, 2007);
- Objetivo: estudar a relação entre o hábito de fumar e mortes por doenças coronarianas;
- Dados: em 1951 foi enviado a todos os médicos britânicos um questionário com questões sobre o hábito de fumar;
- As questões de interesse neste estudo eram:
  - A taxa de mortes é mais alta entre os fumantes?
  - Se sim, quanto ?
  - O efeito diferencial está relacionado à idade ?

Aplicação: Dados da Literatura

- Aplicação (DOBSON, 2007);
- Objetivo: estudar a relação entre o hábito de fumar e mortes por doenças coronarianas;
- Dados: em 1951 foi enviado a todos os médicos britânicos um questionário com questões sobre o hábito de fumar;
- As questões de interesse neste estudo eram:
  - [illegible]
  - O efeito diferencial [illegible]

Aplicação: Dados da Literatura

- Aplicação (DOBSON, 2007);
- Objetivo: estudar a relação entre o hábito de fumar e mortes por doenças coronarianas;
- Dados: em 1951 foi enviado a todos os médicos britânicos um questionário com questões sobre o hábito de fumar;
- As questões de interesse neste estudo eram:
  - A taxa de mortes é mais alta entre os fumantes?
  - Quanto mais alta?
  - O efeito do fumo é o mesmo em homens e mulheres?

Aplicação: Dados da Literatura

- Aplicação (DOBSON, 2007);
- Objetivo: estudar a relação entre o hábito de fumar e mortes por doenças coronarianas;
- Dados: em 1951 foi enviado a todos os médicos britânicos um questionário com questões sobre o hábito de fumar;
- As questões de interesse neste estudo eram:
  - A taxa de mortes é mais alta entre os fumantes?
  - Se sim, quanto ?
  - O efeito diferencial está relacionado à idade ?

Aplicação: Dados da Literatura

- Aplicação (DOBSON, 2007);
- Objetivo: estudar a relação entre o hábito de fumar e mortes por doenças coronarianas;
- Dados: em 1951 foi enviado a todos os médicos britânicos um questionário com questões sobre o hábito de fumar;
- As questões de interesse neste estudo eram:
  - A taxa de mortes é mais alta entre os fumantes?
  - Se sim, quanto ?
  - O efeito diferencial está relacionado à idade ?

Aplicação: Dados da Literatura

- Aplicação (DOBSON, 2007);
- Objetivo: estudar a relação entre o hábito de fumar e mortes por doenças coronarianas;
- Dados: em 1951 foi enviado a todos os médicos britânicos um questionário com questões sobre o hábito de fumar;
- As questões de interesse neste estudo eram:
  - A taxa de mortes é mais alta entre os fumantes?
  - Se sim, quanto ?
  - O efeito diferencial está relacionado à idade ?

Aplicação: Dados da Literatura

- Aplicação (DOBSON, 2007);
- Objetivo: estudar a relação entre o hábito de fumar e mortes por doenças coronarianas;
- Dados: em 1951 foi enviado a todos os médicos britânicos um questionário com questões sobre o hábito de fumar;
- As questões de interesse neste estudo eram:
  - A taxa de mortes é mais alta entre os fumantes?
  - Se sim, quanto ?
  - O efeito diferencial está relacionado à idade ?

Aplicação: Dados da Literatura

Tabela 1. Número de mortes por doenças coronarianas entre os médicos britânicos, após dez anos, categorizados por grupo de idade e hábito de fumar.

Grupo	Fumantes		Não Fumantes	
Idade	Mortes	N. de Pessoas	Mortes	N. de Pessoas
35 – 44	32	52407	2	18790
45 – 54	104	43248	12	10673
55 – 64	206	28612	28	5710
65 – 74	186	12663	28	2585
75 – 85	102	5317	31	1462

Modelo Linear Generalizado

- Para este estudo, um modelo pode ser especificado como:

$$Y_i \sim Poisson(\text{mortes}_i); \qquad E(Y_i) = \text{mortes}_i$$

$$log(\text{mortes}_i) = log(\text{populacao}_i) + \beta_1 + \beta_2 \text{hfumar}_i + \\ + \beta_3 \text{cat.idade}_i + \beta_4 \text{idadesq}_i + \beta_5 \text{hfumidade}_i.$$

  - i denota o i-ésimo grupo definido pelas idades e fumantes ($i = 1, 2, \cdots, 5$) e idades e não fumantes ($i = 6, 7, \cdots, 10$);
  - $mortes_i$ e $populacao_i$ denotam o número esperado e da população no grupo de risco i;
  - $hfumar_i$ é do tipo 0, 1 para não fumantes e fumantes
  - $catidade_i$ toma valores de 1 a 5;
  - $idadesq_i$ raiz quadrada de $ataidade_i$;
  - $hfudade_i$ é o produto de $hfumar_i$ por $catidade_i$.

Modelo Linear Generalizado

- Para este estudo, um modelo pode ser especificado como:

$$Y_i \sim Poisson(\text{mortes}_i); \qquad E(Y_i) = \text{mortes}_i$$

$$log(\text{mortes}_i) = log(\text{populacao}_i) + \beta_1 + \beta_2 \text{hfumar}_i + \\ + \beta_3 \text{cat.idade}_i + \beta_4 \text{idadesq}_i + \beta_5 \text{hfumidade}_i.$$

  - i denota o i-ésimo grupo definido pelas idades e fumantes ($i = 1, 2, \cdots, 5$) e idades e não fumantes ($i = 6, 7, \cdots, 10$);
  - $mortes_i$ e $populacao_i$ denotam o número esperado e da população no grupo de risco i;
  - $hfumar_i$ é do tipo 0, 1 para não fumantes e fumantes
  - $cat.idade$ toma valores de 1 a 5;
  - $idadesq_i$ raiz quadrada de $cat.idade_i$;
  - $hfidade_i$ é o produto de $cat.idade_i$ por $hfumar_i$.

Modelo Linear Generalizado

- Para este estudo, um modelo pode ser especificado como:

$$Y_i \sim Poisson(\text{mortes}_i); \qquad E(Y_i) = \text{mortes}_i$$

$$log(\text{mortes}_i) = log(\text{populacao}_i) + \beta_1 + \beta_2 \text{hfumar}_i + \\ + \beta_3 \text{cat.idade}_i + \beta_4 \text{idadesq}_i + \beta_5 \text{hfumidade}_i.$$

  - i denota o i-ésimo grupo definido pelas idades e fumantes $(i = 1, 2, \cdots, 5)$ e idades e não fumantes $(i = 6, 7, \cdots, 10)$;
  - $mortes_i$ e $populacao_i$ denotam o número esperado e da população no grupo de risco i;
  - $hfumar_i$ é do tipo 0, 1 para não fumantes e fumantes
  - $cat.idade$ toma valores de 1 a 5;
  - $idadeq_i$ raiz quadrada de $cat.idade_i$;
  - $hfidade_i$ é o produto de $cat.idade_i$ por $hfumar_i$.

Modelo Linear Generalizado

- Para este estudo, um modelo pode ser especificado como:

$$Y_i \sim Poisson(\text{mortes}_i); \qquad E(Y_i) = \text{mortes}_i$$

$$log(\text{mortes}_i) = log(\text{populacao}_i) + \beta_1 + \beta_2 \text{hfumar}_i + \\ + \beta_3 \text{cat.idade}_i + \beta_4 \text{idadesq}_i + \beta_5 \text{hfumidade}_i.$$

  - i denota o i-ésimo grupo definido pelas idades e fumantes $(i = 1, 2, \cdots, 5)$ e idades e não fumantes $(i = 6, 7, \cdots, 10)$;
  - $mortes_i$ e $populacao_i$ denotam o número esperado e da população no grupo de risco i;
  - $hfumar_i$ é do tipo 0, 1 para não fumantes e fumantes
  - $cat.idade$ toma valores de 1 a 5;
  - $idadeq_i$ raiz quadrada de $cat.idade_i$;
  - $hfidade_i$ é o produto de $cat.idade_i$ por $hfumar_i$;

Modelo Linear Generalizado

- Para este estudo, um modelo pode ser especificado como:

$$Y_i \sim Poisson(\text{mortes}_i); \qquad E(Y_i) = \text{mortes}_i$$

$$log(\text{mortes}_i) = log(\text{populacao}_i) + \beta_1 + \beta_2 \text{hfumar}_i + \\ + \beta_3 \text{cat.idade}_i + \beta_4 \text{idadesq}_i + \beta_5 \text{hfumidade}_i.$$

  - i denota o i-ésimo grupo definido pelas idades e fumantes ($i = 1, 2, \cdots, 5$) e idades e não fumantes ($i = 6, 7, \cdots, 10$);
  - $mortes_i$ e $populacao_i$ denotam o número esperado e da população no grupo de risco i;
  - $hfumar_i$ é do tipo 0, 1 para não fumantes e fumantes
  - $cat.idade$ toma valores de 1 a 5;
  - $idadeq_i$ raiz quadrada de $cat.idade_i$;
  - $hfidade_i$ é o produto de $cat.idade_i$ por $hfumar_i$;

Modelo Linear Generalizado

- Para este estudo, um modelo pode ser especificado como:

$$Y_i \sim Poisson(\text{mortes}_i); \qquad E(Y_i) = \text{mortes}_i$$

$$log(\text{mortes}_i) = log(\text{populacao}_i) + \beta_1 + \beta_2 \text{hfumar}_i + \\ + \beta_3 \text{cat.idade}_i + \beta_4 \text{idadesq}_i + \beta_5 \text{hfumidade}_i.$$

  - i denota o i-ésimo grupo definido pelas idades e fumantes ($i = 1, 2, \cdots, 5$) e idades e não fumantes ($i = 6, 7, \cdots, 10$);
  - $mortes_i$ e $populacao_i$ denotam o número esperado e da população no grupo de risco i;
  - $hfumar_i$ é do tipo 0, 1 para não fumantes e fumantes
  - $cat.idade$ toma valores de 1 a 5;
  - $idadeq_i$ raiz quadrada de $cat.idade_i$;
  - $hfidade_i$ é o produto de $cat.idade_i$ por $hfumar_i$;

Modelo Linear Generalizado

- Para este estudo, um modelo pode ser especificado como:

$$Y_i \sim Poisson(\text{mortes}_i); \qquad E(Y_i) = \text{mortes}_i$$

$$log(\text{mortes}_i) = log(\text{populacao}_i) + \beta_1 + \beta_2 \text{hfumar}_i + \\ + \beta_3 \text{cat.idade}_i + \beta_4 \text{idadesq}_i + \beta_5 \text{hfumidade}_i.$$

  - i denota o i-ésimo grupo definido pelas idades e fumantes $(i = 1, 2, \cdots, 5)$ e idades e não fumantes $(i = 6, 7, \cdots, 10)$;
  - $mortes_i$ e $populacao_i$ denotam o número esperado e da população no grupo de risco i;
  - $hfumar_i$ é do tipo 0, 1 para não fumantes e fumantes
  - $cat.idade$ toma valores de 1 a 5;
  - $idadeq_i$ raiz quadrada de $cat.idade_i$;
  - $hfidade_i$ é o produto de $cat.idade_i$ por $hfumar_i$;

Aplicação: Dados da Literatura

Tabela 2. Estimativas para os Parâmetros.

Termo	idade.cat	idadeq	fumo	hfidade
$\widehat{\beta}$	$2,376$	$-0,198$	$1,441$	$-0,308$
d.p.$(\widehat{\beta})$	$0,208$	$0,027$	$0,372$	$0,097$
p-valor	$< 0,001$	$< 0,001$	$< 0,001$	$0,002$
R. C.	$10,77$	$0,82$	$4,22$	$0,74$
I.C. 95%	$(7,2; 16,2)$	$(0,78; 0,87)$	$(2,04; 8,76)$	$(0,61; 0,89)$

Aplicação: Dados da Literatura

Tabela 3. Número Observado e Esperado de Óbitos

Categoria de Idade	Hábito de Fumar	Óbitos Observados	Média estimada $\widehat{\mu}$
1	1	32	29,58
2	1	104	106,81
3	1	206	208,20
4	1	186	182,83
5	1	102	102,58
1	0	2	3,54
2	0	12	11,54
3	0	28	27,74
4	0	28	30,23
5	0	31	31,07

Aplicação: Outros Exemplos

Modelos para Dados de Contagem
Modelo de Poisson
Aplicações
Relação entre o hábito de fumar e mortes por doenças coronarianas

Outras Aplicações

- Análise da distribuição espaço temporal da mortalidade pelos cânceres de maior importância quantitativa, segundo as microrregiões do Estado de São Paulo;
- Análise da incidência de malária na Amazonia considerando espaço e tempo;

Aplicação: Outros Exemplos

Modelos para Dados de Contagem
Modelo de Poisson
Aplicações
Relação entre o hábito de fumar e mortes por doenças coronarianas

Outras Aplicações

- Análise da distribuição espaço temporal da mortalidade pelos cânceres de maior importância quantitativa, segundo as microrregiões do Estado de São Paulo;

- Análise da incidência de malária na Amazonia considerando espaço e tempo;

Agradecimentos e Principais Referências

Agradecimentos:

À Comissão organizadora do Evento.

Principais Referências

- Gamerman, D. e Lopes, H. F. (2006) *Markov Chain Monte Carlo: Stochastic Simulation for Baysian Inference*. London: Chapman & Hall/CRC, 2nd. ed.
- Migon, H. S. e Gamerman, D. (1999) it Statistical Inference: an Integrated Approach. Arnold, London, UK.
- Migon, H. S., Souza, A. D. P, Schmidt, A. M. (2008) *Modelos Hierárquicos e Aplicações*. Texto para Minicurso do 18^0_- SINAPE.
  site do livro: www.dme.ufrj.br/ mhierarquico