

# ME610 - Estatística Aplicada

Eliane Siqueira - RA:155233   Guilherme Pazian - RA: 160323

Universidade Estadual de Campinas

12 de maio de 2017

# Sumário

- 1 Objetivo
- 2 Banco de dados
- 3 Preparação dos dados
- 4 Metodologia
- 5 Resultados
- 6 Discussão
- 7 Conclusão

# Banco de dados

Os dados correspondem a informações de um ano de estudo. Mães foram entrevistadas durante a gravidez para coletar informações de outras variáveis que tem influência sobre o peso das crianças.

- 1236 nascimentos
- Crianças nascidas entre 1960 e 1967
- 22 variáveis de interesse

# Banco de dados - Variáveis Categóricas

- Cor\_mae: Raça da mãe (5 níveis)
- Educacao\_mae: Nível educacional da mãe (7 níveis)
- Cor\_pai: Raça do pai (5 níveis)
- Educacao\_pai: Nível educacional do pai (7 níveis)
- Estado\_civil: Estado civil dos pais (5 níveis)
- Rendimento\_anual: Rendimento anual familiar (5 níveis)
- Fuma: A mãe fuma? (4 níveis)
- Tempo\_sem\_fumar: Se a mãe parou de fumar (9 níveis)
- Numero\_cigarros: Número de cigarros fumados por dia (9 níveis)

# Banco de dados - Variáveis Contínuas e discretas

- Peso: Peso do bebê
- Altura\_mae: Altura da mãe
- Peso\_mae: Peso da mãe
- Data\_nasc: Data de nascimento
- Tempo\_gestacao: Duração da gestação
- Numero\_gestacoes: Numero de gestações anteriores
- Idade\_mae: Idade da mãe
- Idade\_pai: Idade do pai

# Banco de dados - Variáveis Descartadas

- Qtd\_feto: Quantidade de fetos
- Vivo: Nascido vivo ou não
- Sexo: Sexo do bebê
- Peso: Peso do bebê
- Altura\_pai: Altura do pai
- Peso\_pai: Peso do pai

# Preparação dos dados

- Observações faltantes
- Observações não classificadas
- Observações absurdas
- Conversão das variáveis Peso e Altura
- 1032 observações restantes

# Metodologia

- Regressão linear múltipla com erros padrão robustos
- Padronização das variáveis contínuas
- Nível de significância  $\alpha = 10\%$
- Reagrupamento dos níveis não significativos ao nível de referência do modelo
- Novo modelo sem as variáveis insignificantes



# Modelo final

$$\begin{aligned} \text{Peso} = & \beta_0 + \beta_1 \text{Fuma} + \beta_2 \text{Mex} + \beta_3 \text{Mixed} + \beta_4 \text{White} \\ & + \beta_5 \left( \frac{P - \bar{P}}{\sigma_P} \right) + \beta_6 \left( \frac{G - \bar{G}}{\sigma_G} \right) + \beta_7 \left( \frac{N - \bar{N}}{\sigma_N} \right) \\ & + \beta_8 \left( \frac{NG - \bar{N}G}{\sigma_{NG}} \right) + \beta_9 \left( \frac{A - \bar{A}}{\sigma_A} \right) + \varepsilon \end{aligned}$$

$\beta_0$ : representa o parâmetro associado aos níveis de referência “Asian/Não definido/Black” (variável Cor Pai) e “Nunca/Fumava” (variável Fuma).

# Resultados

Tabela 1: Ajuste do modelo final

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)( $\beta_0$ )	3.3215	0.0328	101.17	0.0000
FumaFuma( $\beta_1$ )	-0.2390	0.0280	-8.52	0.0000
Cor_paiMex( $\beta_2$ )	0.3447	0.0944	3.65	0.0003
Cor_paiMixed( $\beta_3$ )	0.1886	0.0729	2.59	0.0098
Cor_paiWhite( $\beta_4$ )	0.1847	0.0359	5.15	0.0000
Peso_mae_kg( $\beta_5$ )	0.0428	0.0176	2.43	0.0151
Tempo_gestacao( $\beta_6$ )	0.1970	0.0182	10.85	0.0000
Data_nasc( $\beta_7$ )	0.0282	0.0132	2.13	0.0336
Numero_gestacoes( $\beta_8$ )	0.0495	0.0166	2.99	0.0029
Altura_mae_cm( $\beta_9$ )	0.0701	0.0169	4.14	0.0000

# Discussão

As medidas de poder preditivo foram  $R^2 = 0.2997$  e  $\bar{R}^2 = 0.2936$ .

Variável	Estimativa	$\widehat{\beta}_1$ /Estimativa	EP	$\equiv$ Fumar
<i>Peso_M(kg)</i>	0.0428	-5.58	9.29	-51.88
<i>T_Gestação (dias)</i>	0.1970	-1.21	15.17	-18.35
<i>Num_Gestações</i>	0.0495	-4.83	1.87	-9.03
<i>Altura_M(cm)</i>	0.0701	-3.41	6.47	-22.06

# Conclusão

- Diferenciação clara nas características físicas ou de estilo de vida da mãe.
- As variáveis são intrínsecas à mãe (não estão sujeitas a alterações estipuladas)

# Conclusão

- O fato da mãe fumar durante a gravidez parece ser uma variável mais determinante do peso do bebê do que altura, o peso, a renda anual da mãe e o histórico de resultados de gestações anteriores.