

ME610 - Estatística Aplicada

Eliane e Guilherme

Universidade Estadual de Campinas

26 de maio de 2017

Roteiro

- 1 Objetivo
- 2 Metodologia
- 3 Exemplos
- 4 Resultados
- 5 Discussão
- 6 Conclusão

Objetivo

Primeira pergunta de interesse:

O hábito de fumar influencia no peso do bebê ao nascer?

Segunda pergunta de interesse:

O hábito de fumar tem mais influência no peso do bebê ao nascer, do que altura, peso, renda e histórico de gestações da mãe, além de sexo do bebê e número de fetos ?

- Padronização das variáveis contínuas;

- Padronização das variáveis contínuas;

$$\frac{X - E[X]}{\sqrt{Var(X)}}$$

- Padronização das variáveis contínuas;

$$\frac{X - E[X]}{\sqrt{Var(X)}}$$

- Regressão linear múltipla com erros padrão robustos;

- Padronização das variáveis contínuas;

$$\frac{X - E[X]}{\sqrt{Var(X)}}$$

- Regressão linear múltipla com erros padrão robustos;

- Inferência com base nos testes de nulidade individuais dos parâmetros do modelo;

Metodologia

- Inferência com base nos testes de nulidade individuais dos parâmetros do modelo;
- Nível de significância $\alpha = 10\%$
- Exclusão de variáveis as quais todos os seus níveis foram não significativos;

Metodologia

- Inferência com base nos testes de nulidade individuais dos parâmetros do modelo;
- Nível de significância $\alpha = 10\%$
- Exclusão de variáveis as quais todos os seus níveis foram não significativos;
- Reagrupamento dos níveis não significativos ao nível de referência do modelo;

Metodologia

- Inferência com base nos testes de nulidade individuais dos parâmetros do modelo;
- Nível de significância $\alpha = 10\%$
- Exclusão de variáveis as quais todos os seus níveis foram não significativos;
- Reagrupamento dos níveis não significativos ao nível de referência do modelo;
- Novo modelo sem as variáveis insignificantes;

Exemplo 1: Modelo

$$Y_1 = \mu + \alpha_{a_i} + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i$$

- A é uma variável;
- A_i é o nível i da variável A, $i=1, \dots, 5$.
- $\alpha_{a_1} = 0$ onde α é o incremento na média geral se o indivíduo pertence ao i-ésimo nível da variável A.
- $H_{0i} : \mu_{A_1} = \mu_{A_i}$

Ajuste do Exemplo 1

	Estimativa	Std. Error	t value	Pr(> t)
β_0	3.3030	0.0753	43.86	0.2536
A_1	-0.0021	1.0000	-0.0021	>0.10
A_2	-0.0007	1.0000	-0.0007	>0.10
A_3	0.0086	1.0000	0.0086	>0.10
A_4	-0.0006	1.0000	-0.0006	>0.10
A_5	0.0008	1.0000	0.0008	>0.10
Contínua	16.0000	1.0000	16.0000	<0.0001
Contínua	17.0000	1.0000	17.0000	<0.0004

Ajuste do Exemplo 1

	Estimativa	Std. Error	t value	Pr(> t)
β_0	3.3030	0.0753	43.86	0.2536
A_1	-0.0021	1.0000	-0.0021	≥ 0.10
A_2	-0.0007	1.0000	-0.0007	≥ 0.10
A_3	0.0086	1.0000	0.0086	≥ 0.10
A_4	-0.0006	1.0000	-0.0006	≥ 0.10
A_5	0.0008	1.0000	0.0008	≥ 0.10
Contínua	16.0000	1.0000	16.0000	< 0.0001
Contínua	17.0000	1.0000	17.0000	< 0.0004

Exemplo 2: Modelo

$$Y_i = \mu + \alpha_{a_i} + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i$$

- A é uma variável;
- A_i é o nível i da variável A, $i=1, \dots, 5$.
- $\alpha_{a_1} = 0$ onde α é o incremento na média geral se o indivíduo pertence ao i-ésimo nível da variável A.
- $H_{0i} : \mu_{A_1} = \mu_{A_i}$

Ajuste do Exemplo 2

	Estimativa	Std. Error	t value	Pr(> t)
β_0	3.3030	0.0753	43.86	0.2536
A_1	-0.0021	1.0000	-0.0021	<0.0001
A_2	-0.0007	1.0000	-0.0007	<0.0001
A_3	0.0086	1.0000	0.0086	>0.10000
A_4	-0.0006	1.0000	-0.0006	<0.0001
A_5	0.0008	1.0000	0.0008	<0.0001
Contínua	16.0000	1.0000	16.0000	<0.0001
Contínua	17.0000	1.0000	17.0000	<0.0004

Ajuste de Exemplo 2

	Estimativa	Std. Error	t value	Pr(> t)
β_0	3.3030	0.0753	43.86	0.2536
A_1	-0.0021	1.0000	-0.0021	<0.0001
A_2	-0.0007	1.0000	-0.0007	<0.0001
A_3	0.0086	1.0000	0.0086	>0.10
A_4	-0.0006	1.0000	-0.0006	<0.0001
A_5	0.0008	1.0000	0.0008	<0.0001
Contínua	16.0000	1.0000	16.0000	<0.0001
Contínua	17.0000	1.0000	17.0000	<0.0004

Modelos ajustados

Primeiro modelo

Peso_kg ~ Fuma + Tempo_sem_fumar + Rendimento_anual +
Estado_civil + Educacao_pai + Educacao_mae + Cor_mae + Cor_pai +
Peso_mae_kg + Tempo_gestacao + Data_nasc + Numero_gestacoes +
Idade_mae + Altura_mae_cm + Idade_pai

Variáveis Eliminadas no Primeiro modelo

Peso_kg ~ Fuma + Tempo_sem_fumar + Rendimento_anual +
Estado_civil + Educacao_pai + Educacao_mae + Cor_mae + Cor_pai +
Peso_mae_kg + Tempo_gestacao + Data_nasc + Numero_gestacoes +
Idade_mae + Altura_mae_cm + Idade_pai

Segundo modelo

Peso_kg ~ Fuma + Cor_pai + Cor_mae + Peso_mae_kg +
Tempo_gestacao + Data_nasc + Numero_gestacoes + Altura_mae_cm
(Intercept) : Nunca (Fuma) + Asian (Cor_pai) + Asian (Cor_mae)

Segundo modelo

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	3.3030	0.0753	43.86	0.0000
FumaFuma	-0.2381	0.0304	-7.84	0.0000
FumaFumava	-0.0007	0.0392	-0.02	0.9854
Cor_paiBlack	0.2121	0.1166	1.82	0.0691
Cor_paiMex	0.3512	0.1375	2.55	0.0108
Cor_paiMixed	0.3261	0.1042	3.13	0.0018
Cor_paiNão Definido	0.0603	0.2370	0.25	0.7990
Cor_paiWhite	0.2964	0.0710	4.17	0.0000
Cor_maeBlack	-0.1941	0.1078	-1.80	0.0721
Cor_maeMex	0.0476	0.1149	0.41	0.6787
Cor_maeMixed	-0.0964	0.1141	-0.84	0.3983
Cor_maeNão Definido	-0.0066	0.2010	-0.03	0.9738
Cor_maeWhite	-0.0962	0.0575	-1.67	0.0945
:	:	:	:	:

Reagrupamento dos níveis no Segundo modelo

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	3.3030	0.0753	43.86	0.0000
FumaFuma	-0.2381	0.0304	-7.84	0.0000
FumaFumava	-0.0007	0.0392	-0.02	0.9854
Cor_paiBlack	0.2121	0.1166	1.82	0.0691
Cor_paiMex	0.3512	0.1375	2.55	0.0108
Cor_paiMixed	0.3261	0.1042	3.13	0.0018
Cor_paiNão Definido	0.0603	0.2370	0.25	0.7990
Cor_paiWhite	0.2964	0.0710	4.17	0.0000
Cor_maeBlack	-0.1941	0.1078	-1.80	0.0721
Cor_maeMex	0.0476	0.1149	0.41	0.6787
Cor_maeMixed	-0.0964	0.1141	-0.84	0.3983
Cor_maeNão Definido	-0.0066	0.2010	-0.03	0.9738
Cor_maeWhite	-0.0962	0.0575	-1.67	0.0945
:	:	:	:	:

Terceiro modelo

Peso_kg ~ Fuma + Cor_pai + Cor_mae + Peso_mae_kg +
Tempo_gestacao + Data_nasc + Numero_gestacoes + Altura_mae_cm
(Intercept) : Nunca/Fumava (Fuma) + Asian/Não definido (Cor_pai) +
Asian/Mex/Mixed/Não Definido (Cor_mae)

Terceiro modelo

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	3.3094	0.0691	47.88	0.0000
FumaFuma	-0.2377	0.0283	-8.40	0.0000
Cor_paiBlack	0.1627	0.1123	1.45	0.1477
Cor_paiMex	0.3736	0.1109	3.37	0.0008
Cor_paiMixed	0.2678	0.0994	2.69	0.0072
Cor_paiWhite	0.2711	0.0883	3.07	0.0022
Cor_maeBlack	-0.1519	0.0914	-1.66	0.0971
Cor_maeWhite	-0.0772	0.0634	-1.22	0.2238
Peso_mae_kg	0.0427	0.0179	2.39	0.0171
Tempo_gestacao	0.1975	0.0182	10.83	0.0000
Data_nasc	0.0274	0.0133	2.07	0.0389
Numero_gestacoes	0.0499	0.0166	3.00	0.0027
Altura_mae_cm	0.0710	0.0171	4.16	0.0000

Reagrupamento dos níveis do Terceiro modelo

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	3.3094	0.0691	47.88	0.0000
FumaFuma	-0.2377	0.0283	-8.40	0.0000
Cor_paiBlack	0.1627	0.1123	1.45	0.1477
Cor_paiMex	0.3736	0.1109	3.37	0.0008
Cor_paiMixed	0.2678	0.0994	2.69	0.0072
Cor_paiWhite	0.2711	0.0883	3.07	0.0022
Cor_maeBlack	-0.1519	0.0914	-1.66	0.0971
Cor_maeWhite	-0.0772	0.0634	-1.22	0.2238
Peso_mae_kg	0.0427	0.0179	2.39	0.0171
Tempo_gestacao	0.1975	0.0182	10.83	0.0000
Data_nasc	0.0274	0.0133	2.07	0.0389
Numero_gestacoes	0.0499	0.0166	3.00	0.0027
Altura_mae_cm	0.0710	0.0171	4.16	0.0000

Quarto Modelo

Peso_kg ~ Fuma + Cor_pai + Cor_mae + Peso_mae_kg +
Tempo_gestacao + Data_nasc + Numero_gestacoes + Altura_mae_cm
(Intercept) : Nunca/Fumava (Fuma) + Asian/Não definido/Black
(Cor_pai) + Asian/Mex/Mixed/Não Definido/white (Cor_mae)

Quarto Modelo

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	3.3367	0.0562	59.36	0.0000
FumaFuma	-0.2387	0.0280	-8.52	0.0000
Cor_paiMex	0.3306	0.1027	3.22	0.0013
Cor_paiMixed	0.1781	0.0795	2.24	0.0253
Cor_paiWhite	0.1694	0.0570	2.97	0.0030
Cor_maeBlack	-0.0197	0.0635	-0.31	0.7559
Peso_mae_kg	0.0434	0.0177	2.46	0.0142
Tempo_gestacao	0.1969	0.0182	10.83	0.0000
Data_nasc	0.0282	0.0132	2.13	0.0333
Numero_gestacoes	0.0499	0.0167	2.99	0.0029
Altura_mae_cm	0.0706	0.0170	4.15	0.0000

Reagrupamento dos níveis do Quarto modelo

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	3.3367	0.0562	59.36	0.0000
FumaFuma	-0.2387	0.0280	-8.52	0.0000
Cor_paiMex	0.3306	0.1027	3.22	0.0013
Cor_paiMixed	0.1781	0.0795	2.24	0.0253
Cor_paiWhite	0.1694	0.0570	2.97	0.0030
Cor_maeBlack	-0.0197	0.0635	-0.31	0.7559
Peso_mae_kg	0.0434	0.0177	2.46	0.0142
Tempo_gestacao	0.1969	0.0182	10.83	0.0000
Data_nasc	0.0282	0.0132	2.13	0.0333
Numero_gestacoes	0.0499	0.0167	2.99	0.0029
Altura_mae_cm	0.0706	0.0170	4.15	0.0000

Variáveis Eliminadas no Quarto Modelo

Peso_kg ~ Fuma + Cor_pai + Cor_mae + Peso_mae_kg +
Tempo_gestacao + Data_nasc + Numero_gestacoes + Altura_mae_cm
(Intercept) : Nunca/Fumava (Fuma) + Asian/Não definido/Black
(Cor_pai)

Modelo final

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

β_0 : representa o parâmetro associado aos níveis de referência “Asian/Não definido/Black” (variável Cor Pai) e “Nunca/Fumava” (variável Fuma).

P, G, N, NG, A: são os valores observados para as variáveis, “Peso_mae_kg” “Tempo_gestacao”, “Data_nasc”, “Numero_gestacoes” e “Altura_mae_cm”, respectivamente

Ajuste do modelo final

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)(β_0)	3.3215	0.0328	101.17	0.0000
FumaFuma(β_1)	-0.2390	0.0280	-8.52	0.0000
Cor_paiMex(β_2)	0.3447	0.0944	3.65	0.0003
Cor_paiMixed(β_3)	0.1886	0.0729	2.59	0.0098
Cor_paiWhite(β_4)	0.1847	0.0359	5.15	0.0000
Peso_mae_kg(β_5)	0.0428	0.0176	2.43	0.0151
Tempo_gestacao(β_6)	0.1970	0.0182	10.85	0.0000
Data_nasc(β_7)	0.0282	0.0132	2.13	0.0336
Numero_gestacoes(β_8)	0.0495	0.0166	2.99	0.0029
Altura_mae_cm(β_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

Primeira pergunta

O hábito de fumar tem influência no peso da criança ao nascer.

Segunda pergunta

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. Não podemos tirar conclusões a partir desses dados, para o sexo do bebê e número de fetos, pois ambas eram constantes no banco de dados disponível.