Universidade Estadual de Campinas

Efeito do Fumo no Peso Médio de Bebês ao Nascer

Trabalho Final

Nome :	Flávio Augusto Zamot Ferreira, Responsável	**RA** : 070913
	Henrique de Souza e Silva	169595
	Leonardo da Silva Araújo	220044
	Vinícius Prado Moreira Silva	195079

Campinas
Novembro de 2023

Resumo: O objetivo deste trabalho é averiguar quão significativo é o efeito do hábito de fumar durante toda a gravidez sobre o peso médio de bebês ao nascerem quando comparado ao efeito de outros condicionantes do crescimento dos fetos. Para tanto, é ajustado um modelo de Regressão Linear Múltipla aos dados acerca de 1.236 nascimentos ocorridos no *Kaiser Foundation Hospital* em Oakland, California na década de 1960. Conforme os resultados do modelo escolhido, o efeito do hábito de fumar durante toda a gravidez mostrou-se mais significativo que o efeito do peso, do número de gestações anteriores, da renda e do grau de instrução da mãe.

Palavras-chave: tabagismo; gestação; regressão linear.

Conteúdo

Lista de Figuras

Lista de Tabelas

1 Introdução

Conforme Franceschini et al. (2003), o peso de um bebê ao nascer é condicionado pela atenção dispensada à gestante, bem como a seu estado nutricional tanto antes, quanto durante o período gestacional e a sua exposição a fatores de risco. Dentre esses fatores, pode-se citar o tabagismo.

Embora, na atualidade, os malefícios do cigarro sejam de conhecimento público, o reconhecimento dos riscos que o fumo traz à saúde dos usuários veio muito depois do estabelecimento e da internacionalização da indústria do cigarro. Em verdade, a atitude da opinião pública em relação ao cigarro durante grande parte do século XX era bastante mais favorável do que é atualmente.

Conforme faz o próprio sítio oficial da agência pública *Centers for Disease Control and Prevention*, não é incomum citar a publicação em 1964 do primeiro relatório do *Surgeon General's Advisory Committee on Smoking and Health* como marco fundamental da consolidação das evidências científicas existentes — e, portanto, da reversão do processo de difusão do tabagismo.

O objetivo deste trabalho é averiguar quão significativo é o efeito do hábito de fumar durante toda a gravidez sobre o peso médio de bebês ao nascerem quando comparado ao efeito de outros condicionantes do crescimento dos fetos.

Para tanto, dispõe-se de dados acerca de 1.236 nascimentos ocorridos no *Kaiser Foundation Hospital* em Oakland, California nos anos de 1961 e 1962, aos quais optou-se aplicar métodos de análise de Regressão Linear Múltipla. Note que, à época, os malefícios do cigarro ainda estavam longe de ser consenso.

Este trabalho está estruturado em seis seções além desta introdução. Na segunda seção, explica-se, em mais detalhes, os dados de que se dispõe e o tratamento a eles dados. Em seguida, discorre-se acerca dos resultados do modelo de Regressão Linear Múltipla escolhido. Na quarta seção, algumas conclusões são apresentadas. Na quinta seção, as responsabilidades de cada membro do grupo são explicitadas. Em seguida, listam-se as referências; e, por fim, disponibiliza-se *links* para o acesso aos códigos utilizados para a confecção deste documento.

2 Dados

Para cumprir os objetivos deste trabalho, dispõe-se de dados provenientes do *Child Health and Development Studies* (CHDS) fornecidos pelo cliente. Trata-se de um estudo que inclui informações sobre as crianças nascidas entre 1960 e 1967 no *Kaiser Foundation Hospital* em Oakland, Califórnia.

No banco de dados de que foi disponibilizado, registram-se informações acerca de 1.236 nascimentos de bebês que, nos anos de 1961 e de 1962, sobreviveram, pelo menos, 28 dias. Para cada um dos nascimentos, dispõe-se de 23 variáveis, dentre as quais se encontra o peso do bebê ao nascer `wt`.

Das 23 variáveis presentes no banco de dados original, algumas foram desconsideradas de antemão. As variáveis `pluralty` —, isto é, o número de fetos em gestação —, `outcome` —, isto é, o resultado da gestação — e `sex` —, isto é, o sexo do bebê — foram ignoradas porque consta, apenas, um valor para cada no banco de dados. Note, então, que o conjunto de dados consiste em informações acerca de 1.236 gestações de feto único do sexo masculino que resultaram no nascimento de um bebê que sobreviveu, ao menos, 28 dias.

Ademais, a variável `date` —, isto é, a data do nascimento — foi descartada pela complexidade que introduziria na análise. As variáveis `race` e `drace` —, ou seja, a raça da mãe e do pai respectivamente — foram ignoradas por razões éticas. Descartou-se, também, o estado civil da mãe — reportado na variável `marital` — por sua irrelevância presumida tendo em mente as demais preditoras presentes no banco de dados.

Por fim, decidiu-se desconsiderar a variável `time`. Como se optou por avaliar o efeito de a mãe fumar durante toda a gestação, o período desde o momento em que a mãe ex-fumante parou de fumar não é relevante para a análise.

Assim, o banco de dados de trabalho contém, além do número de identificação da observação `id`, 14 variáveis.

- `gestation`: a duração da gestação em dias;
- `bwt`: o peso do bebê ao nascer em onças;
- `parity`: o número de gestações — mesmo que interrompidas — anteriores da mãe;
- `mage`: a idade da mãe em anos ao fim da gestação;
- `med`: o grau de instrução da mãe — em que 0 corresponde a ter cursado até menos do que a 8ª série; 1 a ter cursado até, no máximo, o 3º ano do Ensino Médio sem formar-se; 2, a ter interrompido os estudos após formar-se no Ensino Médio; 3, a ter, além do Ensino Médio, um diploma de Ensino Técnico; 4, a ter cursado algum tempo de Ensino Superior sem formar-se; 5, a ter o Ensino Superior Completo; 6 e 7, a tempo incerto de Ensino Médio ou Técnico;
- `mht`: a altura da mãe em polegadas;
- `mwt`: o peso da mãe em libras durante a gravidez;
- `dage`: a idade do pai em anos;
- `ded`: o grau de instrução do pai com codificação similar ao da mãe;
- `dht`: a altura do pai em polegadas;
- `dwt`: o peso do pai em libras;
- `inc`: renda familiar anual em incrementos de 2.500 dólares; em que 0 corresponde a uma renda menor do que 2.500 dólares, e 9, a uma renda maior do que 22.500 dólares;
- `smoke`: o hábito de fumo da mãe — em que 0 corresponde a nunca ter fumado; 1, a fumar no momento da entrevista; 2, a fumar até o início da gravidez atual; 3, a ter fumado em algum momento do passado; 9, a uma situação desconhecida;
- `number`: o número de cigarros fumados pela mãe seja atualmente, seja no passado — em que 0 corresponde a nunca ter fumado; 1, a fumar de 1 a 4 cigarros; 2, a de 5 a 9; 3, a de 10 a 14; 4, a de 15 a 19; 5, a de 20 a 29; 6, a de 30 a 39; 7, a de 40 a 60; 8, a mais do que 60; 9, a não saber precisar o número; 98, a um número desconhecido; 99, a não ter sido perguntado à mãe.

Note que algumas variáveis foram renomeadas — como, por exemplo, o peso da mãe e o grau de instrução da mãe — para evitar ambiguidades com as informações correspondentes seja ao bebê, seja ao pai.

Realizou-se algumas redefinições dos níveis das variáveis qualitativas. Tanto `med`, quanto `ded` foram recodificadas com vistas a terem, apenas, dois níveis: nenhum período de Ensino Superior e algum tempo de Ensino Superior. Decidiu-se, por esse critério de separação, porque o efeito de ter algum tempo de Ensino Superior é, justamente, o efeito em favor do qual as evidências são mais favoráveis.

Além disso, optou-se por considerar os níveis 6 e 7 — correspondentes a completude de Ensino Médio ou Técnico incerta — como dados faltantes no que tange ao grau de instrução tanto da mãe, quanto do pai.

O número de cigarros fumados por dia `number`, por sua vez, foi recodificado para dispor de, apenas, três níveis: 0 para corresponder a não ter fumado durante toda a gravidez; 1, a fumar de 1 a 14 cigarros por dia durante toda a gravidez; e 2, a fumar mais de 14 cigarros por dia durante toda a gravidez.

Ademais, note que, se a mãe declarou ter largado o cigarro em algum momento anterior ao nascimento do bebê, foi atribuído o nível 0 da variável `number` recodificada. O motivo, para isso, é a escolha por isolar o efeito do hábito de fumar consistente ao longo de toda a gravidez.

Para as análises realizadas, optou-se por trabalhar, apenas, com casos completos das variáveis de interesse. Assim sendo, as observações em que as informações correspondentes a, ao menos, uma das

variáveis envolvidas — seja a resposta, seja uma das preditoras — foram identificadas como faltantes foram descartadas.

Por fim, cabe informar que, para o tratamento do banco de dados, foram utilizadas as funções da biblioteca `tidyverse` em sua versão 2.0.0. Para a identificação das observações faltantes tal como codificadas no estado inicial do conjunto de dados, foi utilizada a função `set_na()` tal como implementada na versão 1.2.0 do pacote `sjlabelled`.

3 Métodos

Primeiramente, é importante alertar que, a não ser quando relatado o contrário, os resultados desta seção foram obtidos por meio de funções disponíveis na biblioteca básica da versão 4.3.1 do `R`. Todos os gráficos foram feitos utilizando o pacote `ggplot2` em sua versão 3.4.3.

Tabela 1: Estatísticas sumárias de variáveis selecionadas

Variável	Mínimo	1º Quartil	Média	Mediana	3º Quartil	Desvio	N. Faltantes
`bwt`	55	108,75	119,58	120	131	18,24	0
`gestation`	148	272,00	279,34	280	288	16,03	13
`parity`	0	0,00	1,93	1	3	1,93	0
`mht`	53	62,00	64,05	64	66	2,53	22
`dwt`	110	155,00	171,20	170	185	22,39	499

A **Tabela 1** traz algumas estatísticas sumárias de algumas das variáveis incluídas no modelo escolhido. Note que o número de observações faltantes do peso do pai `dwt` é maior que o das demais.

É importante reconhecer, assim, que a inclusão dessa variável aumenta consideravelmente o número de casos incompletos. Então, o número de casos incluídos no modelo final — qual seja, 711 — é bem menor que o número total de observações contidas no banco de dados original.

Como a varíavel `number` recodificada tem apenas três níveis, cabe reportar, brevemente, o número de observações em cada um deles. 742 mães alegaram não terem fumado durante todo o período gestacional enquanto que 261 admitiram entre 1 e 14 cigarros por dia. Ademais, 219 reconheceram que fumaram, pelo menos, 15 cigarros por dia durante a gestação. Por fim, são 14 as observações faltantes da variável `number` recodificada.

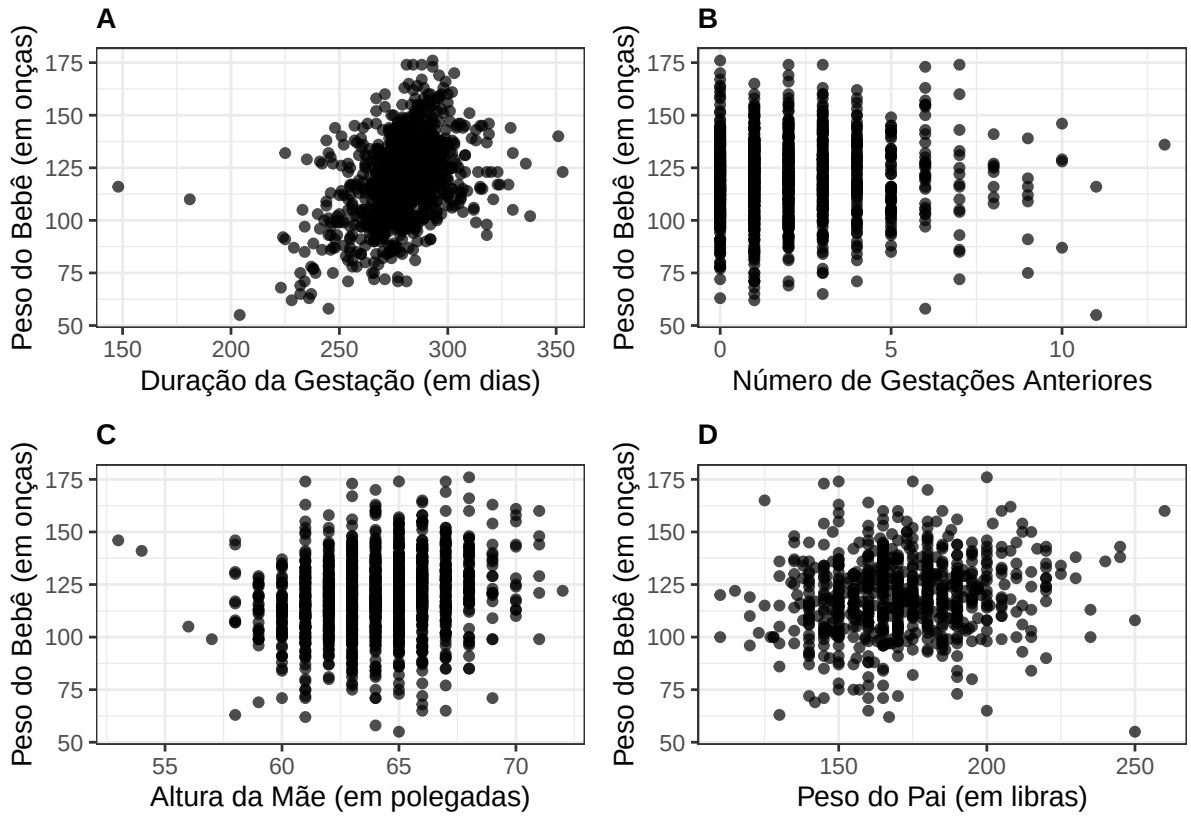

Figura 1: Gráfico de Dispersão do Peso do Bebê ao Nascer contra **A**: a Duração da Gestação; **B**: o Número de Gestações Anteriores da Mãe; **C**: a Altura da Mãe; **D**: o Peso do Pai

Os gráficos de dispersão da **Figura 1** permitem avaliar, preliminarmente, a associação entre a variável resposta e algumas das variáveis preditoras selecionadas.

Em primeiro lugar, parece haver indícios, no painel **A** da **Figura 1**, de uma associação positiva forte entre o peso do bebê ao nascer e a duração de gestação. Trata-se de um resultado esperado tendo em mente conhecimento básico de desenvolvimento intra-uterino. Ademais, dois pontos amostrais aparecem desgarrados à esquerda da nuvem de observações.

Os paineis **C** e **D** da **Figura 1** parecem, por sua vez, trazer indícios de alguma associação positiva do peso do bebê tanto com a altura da mãe, quanto com o peso do pai. Por outro lado, parece difícil distinguir alguma tendência no gráfico de dispersão de `bwt` contra o número de gestações anteriores da mãe conforme o painel **B** da **Figura 1**.

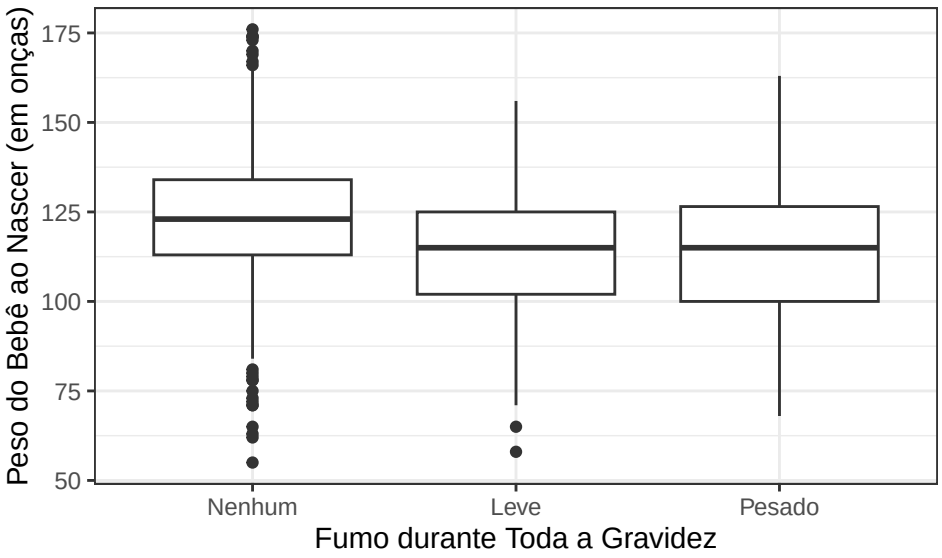

Figura 2: Diagrama de Caixa do Peso do Bebê ao Nascer contra o Número de Cigarros Fumados pela Mãe por Dia.

O diagrama de Caixa reproduzido na **Figura 2** permite, por fim, uma primeira apreciação dos indícios acerca da relação entre o peso do bebê ao nascer e o hábito de a mãe fumar durante toda a gravidez. Note que as caixas associadas às mães que declararam fumar, ao menos, um cigarro por dia ao longo de toda a gravidez aparecem rebaixadas quando comparadas à caixa correspondente às mães que alegaram não terem fumado, consistentemente, durante a gravidez.

Cabe uma explicação breve sobre o procedimento de seleção de modelo empregado. A função `regsubsets()` tal como implementada na versão 3.1 do pacote `leaps` seleciona, para cada número de preditoras incluídas no modelo de Regressão Linear Múltipla, o melhor subconjunto de variáveis explicativas.

Com base nessa informação, é possível escolher o número de preditoras incluídas no modelo — e, por tanto, o melhor subconjunto de variáveis explicativas — que maximiza o coeficiente de determinação ajustado. Assim sendo, foi escolhido, inicialmente, um modelo de Regressão Linear Múltipla para a variável resposta `bwt` que incluísse `gestation`, `parity`, `mht`, `ded`, `dwt` e `number`.

Um ajuste preliminar deste modelo permite observar que a força das evidências favoráveis à hipótese de não nulidade dos coeficientes de regressão associados ao grau de instrução do pai recodificado e à renda familiar em incrementos de 2.500 dólares é, consideravelmente, menor que a do coeficiente correspondente às demais variáveis preditoras incluídas. Decidiu-se, assim, retirá-las.

A fim de facilitar a comparação dos efeitos, optou-se por padronizar a resposta `bwt` e as preditoras `mht` e `dwt`. Que uma estrela $\star$ simbolize que a variável em questão foi padronizada. Então, o modelo final escolhido é, para cada caso completo i de gestação de feto único do sexo masculino que resultou no nascimento de um bebê que sobreviveu, ao menos, 28 dias,

$$\texttt{bwt}_i^\star = \beta_0 + \beta_1\,\texttt{gestation}_i + \beta_2\,\texttt{parity}_i + \beta_3\,\texttt{mht}_i^\star + \beta_4\,\texttt{dwt}_i^\star + \beta_5\,\texttt{number}_i + \varepsilon_i$$

em que

- β_j é fixo e desconhecido para todo $j \in \{0, 1, \cdots, 5\}$;
- $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$ com $\sigma^2 > 0$ fixo e conhecido;
- $\text{Cov}(\varepsilon_i, \varepsilon_j) = 0$ para todo $j \neq i$.

Antes de mais nada, é imperativo averiguar a razoabilidade das hipóteses do modelo proposto.

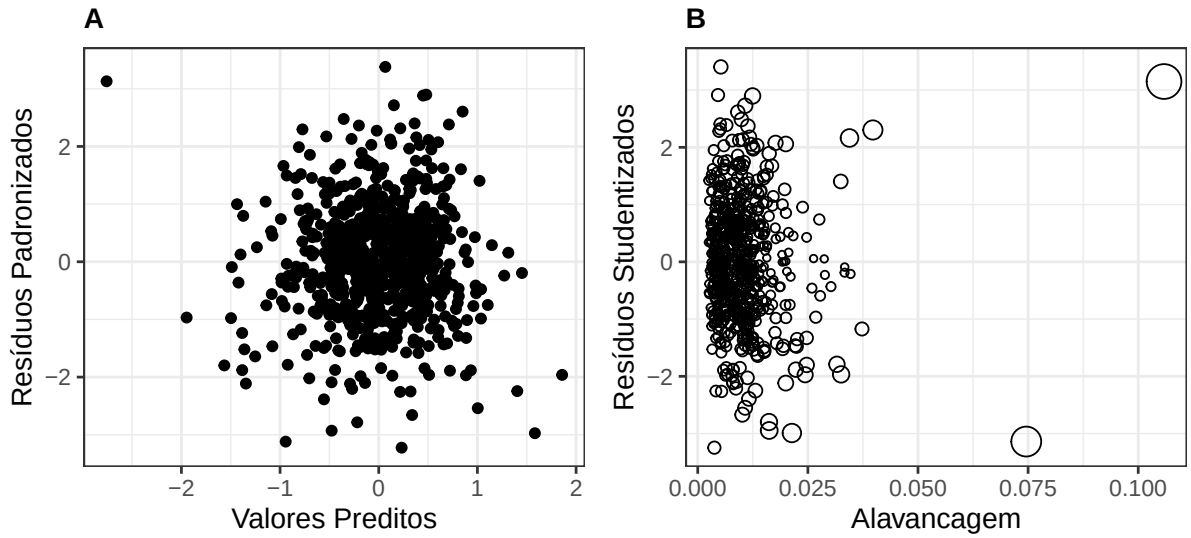

Figura 3: **A**: Gráfico de Dispersão dos Resíduos Padronizados contra os Valores Preditos. **B**: Gráfico de Dispersão dos Resíduos Studentizados contra a Alavancagem. O tamanho dos pontos representa a Distância de Cook.

No gráfico de dispersão reproduzido no painel **A** da **Figura 3**, não parece ser discernível relação alguma entre os resíduos padronizados e os valores preditos. Note, também, que a dispersão dos resíduos não parece variar de forma sistemática com os valores preditos. Preocupam, no entanto, resíduos potencialmente discrepantes localizados no canto inferior direito e, principalmente, um resíduo localizado no canto superior esquerdo do gráfico.

Parece ser oportuno avaliar a influência de potenciais *outliers* no resultado de Regressão. O gráfico reproduzido no painel **B** da **Figura 3** permite, em alguma medida, fazer isso. Note que se desgarram da nuvem de pontos dois resíduos studentizados grandes em módulo de alavancagem maior que a dos demais, o que resulta em uma Distância de Cook também consideravelmente maior. Assim sendo, a presença de casos influentes, se não remediada a contento, pode colocar em dúvida a razoabilidade do modelo proposto.

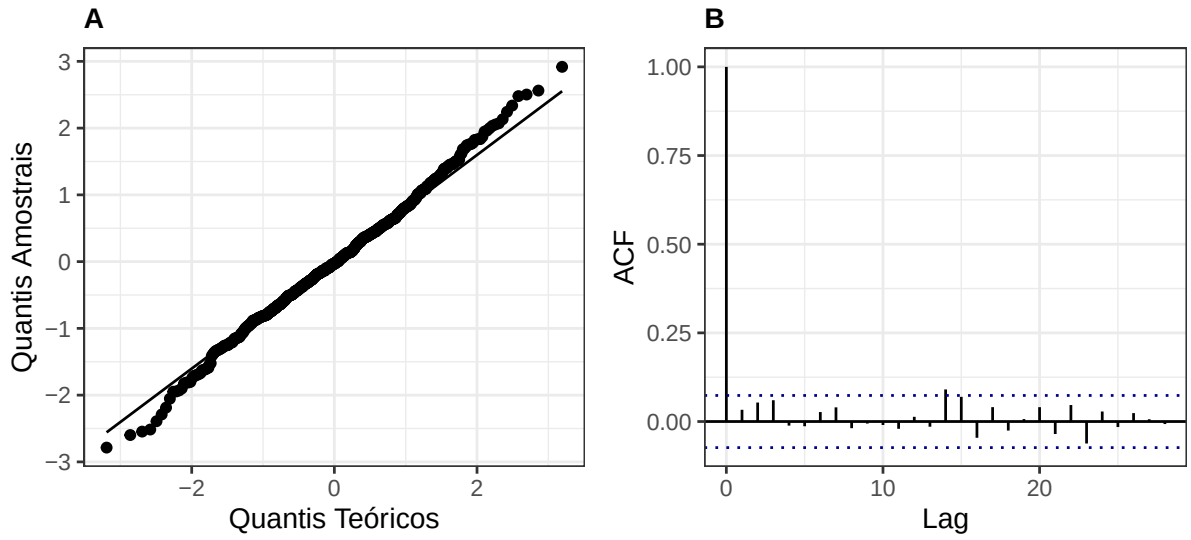

Figura 4: **A**: Gráfico de Probabilidade Normal dos Resíduos. **B**: Gráfico de Autocorrelação dos Resíduos.

Conforme o gráfico de probabilidade normal reproduzido no painel **A** da **Figura 4**, a hipótese de normalidade dos erros parece razoável. Note que, apesar de um descolamento leve discernível nos extremos do gráfico, os pontos parecem bem aglomerados ao redor de uma linha. O gráfico de autocorrelação apresentado no painel **B** da **Figura 4** não traz indícios de irrazoabilidade da hipótese de independência dos erros.

A análise dos gráficos apropriados para confrontar os resíduos em valor absoluto e as preditoras incluídas no modelo não apresentam grandes indícios de que a hipótese de homocedasticidade não seja razoável.

A análise dos Fatores de Inflacionamento de Variância não traz indícios de que o modelo de Regressão Múltipla escolhido sofra com problemas de multicolinearidade. De fato, tanto o VIF médio, quanto o VIF máximo estão bem próximos da unidade.

Remediada a presença de casos influentes, pode-se seguir para a análise dos resultados do ajuste do modelo de Regressão Linear Múltipla escolhido.

Tabela 2: Resultados do Modelo de Regressão Linear Múltipla Ajustado conforme Método de Estimação

Termo	Estimativa	Est. Huber[1]	Est. Tukey[2]	Estatística	Estat. Huber[1]	Estat. Tukey[2]
Intercepto	-6,65019	-6,66287	-6,72215	-11,43	-11,51	-11,59
`gestation`	0,02418	0,02424	0,02444	11,75	11,84	11,92
`parity`	0,03375	0,03028	0,03174	1,99	1,80	1,88
`mht`*	0,18796	0,18519	0,18387	5,60	5,55	5,50
`dwt`*	0,09427	0,10605	0,11399	2,79	3,16	3,39
Fumo Leve	-0,47553	-0,49338	-0,48736	-5,63	-5,87	-5,79
Fumo Pesado	-0,47748	-0,48404	-0,48516	-5,31	-5,41	-5,41

Os resultados referem-se a um modelo de regressão linear múltipla em que a resposta e algumas preditoras se encontram padronizadas.

[1] As estimativas robustas referem-se ao resultado de uma estimação-M para um modelo de Regressão Linear Iterativamente Ponderada usando a função de pesos de Huber.

[2] As estimativas robustas referem-se ao resultado de uma estimação-MM para um modelo de Regressão Linear Iterativamente Ponderada usando a função de pesos biquadrados de Tukey.

A **Tabela 2** permite comparar o valor da estimativa de Mínimos Quadrados Ordinários e da estatística associada a duas alternativas robustas. Para o ajuste da Regressão Robusta comparável, foi utilizada a função `rlm()` do pacote `MASS` em sua versão 7.3.60.

Note que tanto as estimativas, quanto as estatísticas de teste associados aos coeficientes de regressão não se modificam muito quando se procura remediar a influência de alguns casos. Não parece haver, portanto, grande prejuízo em seguir a análise assumindo que o modelo de Regressão Linear Múltipla é razoável.

No entanto, é importante reconhecer que o valor absoluto da estatística robusta de teste do coeficiente de regressão associado ao número de gestações anteriores da mãe `parity` é, consideravelmente, menor para ambos os métodos. Isso é relevante porque se trata, justamente, da menor estatística de teste em módulo.

Tabela 3: Resultados do Modelo de Regressão Linear Múltipla Ajustado por Mínimos Quadrados Ordinários

Termo	Estimativa	Erro Padrão	Valor de p
Intercepto	-6,65019	0,58169	$7{,}0 \times 10^{-28}$
`gestation`	0,02418	0,00206	$3{,}0 \times 10^{-29}$
`parity`	0,03375	0,01693	$4{,}7 \times 10^{-2}$
`mht`*	0,18796	0,03356	$3{,}1 \times 10^{-8}$
`dwt`*	0,09427	0,03375	$5{,}4 \times 10^{-3}$
Fumo Leve	-0,47553	0,08447	$2{,}6 \times 10^{-8}$
Fumo Pesado	-0,47748	0,08994	$1{,}5 \times 10^{-7}$

Os resultados referem-se a um modelo de regressão linear múltipla em que a resposta e algumas preditoras se encontram padronizadas. Para os valores-de-p, assume-se que as estatísticas têm distribuição t com 711 graus de liberdade.

A **Tabela 3** traz os resultados usuais do ajuste do modelo de Regressão Linear Múltipla por Mínimos Quadrados Ordinários. O coeficiente de determinação múltipla ajustado do modelo escolhido é, aproximadamente, $R^2_{\text{adj}} \approx 0,257$.

Note que, conforme mencionado anteriormente, a força das evidências favoráveis à não nulidade do coeficiente de regressão associado à `parity` é menor do que a força das evidências favoráveis aos coeficientes das demais preditoras incluídas. Como os resultados robustos parecem sugerir que a força dessas evidências é, remediada a influências de alguns casos destoantes, ainda menor, pode lançar-se dúvida acerca sobre a significância do coeficiente associado ao número de gestações anteriores da mãe.

Na **Tabela 3**, ainda chama a atenção a magnitude da estimativa pontual do efeito do hábito de fumar da mãe sobre o peso do bebê ao nascer. Tudo o mais constante, estima-se que o peso médio do bebê nascer tende a ser cerca de meio desvio padrão —, ou seja, $8,71$ onças aproximadamente — se a mãe manteve o hábito de fumar durante toda a gravidez.

Trata-se de uma redução considerável tendo em mente que o peso médio ao nascer dos bebês no conjunto de dados utilizado no ajuste é cerca de $119,5$ onças. Por outro lado, note que, conforme os resultados do ajuste, o efeito de fumar entre 1 e 14 cigarros por dia não é distinguível que fumar mais que 14 cigarros por dia.

De fato, a força da evidência favorável à não nulidade do coeficiente de regressão associado ao hábito de fumar — seja leve, seja pesado — da mãe durante toda a gravidez tal como medida pelo valor de p só não é maior que a força das evidências favoráveis ao efeito do número total de dias de gestação `gestation` e ao efeito da altura da mãe `mht`. Note que o valor de p associado ao efeito de fumar entre 1 e 14 cigarros por dia durante toda a gravidez é da mesma ordem de magnitude que o valor de p correspondente ao coeficiente associado a `mht`, mas um pouco menor ainda assim.

Portanto, parece haver evidências suficientes para afirmar que o hábito de a mãe fumar cigarros durante toda a gravidez está associado, tudo o mais constante, a um peso médio do bebê ao nascer menor, pelo menos, para gestações de fetos do sexo masculino levadas a termo na região atendida pelo *Kaiser Foundation Hospital* em Oakland, California entre 1961 e 1962.

4 Conclusão

Ajustou-se um modelo de Regressão Linear Múltipla a dados acerca de 1.236 gestações de feto único do sexo masculino que, no *Kaiser Foundation Hospital* em Oakland, California entre os anos de 1961 e de 1962, resultaram no nascimento de um bebê que sobreviveu, ao menos, 28 dias.

Das 23 variáveis disponíveis no banco de dados originais, 8 foram ignoradas de antemão. Das 15 restantes, outras 8, além do número de identificação das observações, foram desconsideradas conforme os critérios de seleção de modelo empregados.

Como se identificou a presença de alguns dados influentes, métodos de Regressão Robusta foram aplicados. Uma vez que a estimativa robusta dos coeficientes de regressão e a estatística robusta de teste associada não foram, radicalmente, diferentes dos resultados de Mínimos Quadrados Ordinários, considerou-se razoável o modelo proposto.

Com base nesse modelo, estima-se que, tudo o mais constante, o peso médio do bebê nascer tende a ser cerca de $8,71$ onças menor se a mãe manteve o hábito de fumar durante toda a gravidez. Embora não seja muito adequado comparar a magnitude de coeficientes de regressão associados a variáveis quantitativas e qualitativas, chama a atenção o tamanho do efeito do hábito de a mãe fumar durante toda a gravidez tendo em mente que o peso médio ao nascer dos bebês presentes no banco de dados utilizado para o ajuste é $119,5$ onças.

De fato, os coeficientes de regressão associados ao número de cigarros fumados pela mãe por dia ao longo de toda a gravidez são, conforme os resultados do modelo ajustado, mais significativos que os coeficientes correspondentes ao número de gestações anteriores da mãe e ao peso do pai. Ademais, o coeficiente de regressão associado a fumar entre 1 e 14 cigarros por dia durante a gestação também é um pouco mais significativo que aquele correspondente à altura da mãe.

Note que, embora o peso, o grau de instrução da mãe e a renda familiar não estejam inclusas no modelo escolhido, é possível afirmar que, conforme a natureza do procedimento de seleção de modelo empregado, os coeficientes de regressão associados a essas variáveis são, nesse conjunto de dados, menos significativos que aqueles associados ao número de cigarros fumados pela mãe por dia ao longo de toda a gravidez.

Devido à limitação dos dados disponíveis na amostra, não é possível, no entanto, comparar a significância do hábito de fumar durante a gravidez à significância do sexo do bebê ou do resultado de gestações anteriores para o peso do bebê ao nascer.

Assim sendo, conclui-se que há, no banco de dados analisados, indícios de que o resultado reportado no relatório do *US Surgeon General* de 1989

> "[...] cigarrette smoking seems to be a more significant determinant of birthweight than the mother's pregnancy height, weight, parity, payment status, or history of previous pregnancy outcome, or the infant's sex. The reduction in birthweight associated with maternal tobacco use seems to be a direct effect of smoking on fetal growth." (US Department of Health and Human Services, 1989, p. 71).

esteja —, pelo menos, em partes — correto.

5 Responsabilidades

Banco de dados: Flávio

Modelagem: Flávio e Henrique

Redação: Flávio, Henrique, Leonardo e Vinícius

6 Referências

FOX, J.; WEISBERG, S.. **An R Companion to Applied Regression**. 3 ed.. Thousand Oaks: Sage, 2019.

FRANCESCHINI, S. do C. C. et al.. Fatores de risco para o baixo peso ao nascer em gestantes de baixa renda. **Revista De Nutrição**, v. 16, n. 2, pp. 171–179, 2003.

LÜDECKE, D. "sjlabelled: Labelled Data Utility Functions". **The Comprehensive R Archive Network**, 2020. Disponível em: https://CRAN.R-project.org/package=sjlabelled. Acesso em 21 de nov. de 2023.

KUTNER, M. H. et. al..**Applied Linear Statistical Models**. 5 ed.. Nova Iorque: McGraw-Hill Irwin, 2004.

LUMLEY, T. "leaps: Regression Subset Selection". **The Comprehensive R Archive Network**, 2020. Disponível em: https://CRAN.R-project.org/package=leaps. Acesso em 21 de nov. de 2023.

R CORE TEAM. **R**: a Language and Environment for Statistical Computing. Vienna: R Foundation for Statistical Computing, 2023.

US DEPARTMENT OF HEALTH AND HUMAN SERVICES. *Reducing the Health Consequences of Smoking*: 25 Years of Progress. A Report of the Surgeon General. DHHS Publication n. (CDC) 89-8411, 1989.

VENABLES, W.N.; RIPLEY, B.D..**Modern Applied Statistics with S**. 4 ed.. Nova Iorque: Springer, 2002.

WICKHAM, H.. **ggplot2**: Elegant Graphics for Data Analysis. Nova Iorque: Springer-Verlag, 2016.

WICKHAM, et. al.. "Welcome to the tidyverse". **Journal of Open Source Software**, v. 4, n. 43, p. 1686, 2019. Disponível em: https://joss.theoj.org/papers/10.21105/joss.01686. Acesso em: 21 de nov. de 2023.

7 Apêndice

O *script* contendo o código utilizado para o tratamento dos dados e o ajuste do modelo de Regressão Múltipla escolhido está disponível **neste link**. Ademais, os códigos necessários para reproduzir tanto as tabelas, quanto as figuras reproduzidas podem ser consultados diretamente no arquivo `RMarkdown` utilizado para gerar este documento acessível a partir **deste link**. São necessárias credenciais da Universidade Estadual de Campinas para acessá-los.