

Uso do modelo multinomial na identificação de fatores associados à obesidade

L. B. Nucci¹, A. C. C. N. Mafra², L. T. O. Zangirolani³, R. Cordeiro⁴
DMPS, FCM – UNICAMP, 13.083-887, Campinas/SP.

L. Y. Barros⁵
Depto de Estatística, IMECC – UNICAMP, 13.083-859, Campinas/SP.

Resumo. Este trabalho tem como objetivo aplicar o modelo de regressão multinomial para a identificação de fatores de risco relacionados à obesidade no Distrito Sul de Campinas, interior de São Paulo. A associação do excesso de peso a problemas cardiovasculares, articulares e até mesmo sociais mostram a importância de também se identificar fatores que influenciam a saúde devido à condição física do indivíduo. Foram propostos dois modelos que procuram descrever, separadamente, a relação entre hábitos alimentares, medidas antropométricas e variáveis socioeconômicas e demográficas com risco de obesidade e de doenças cardiovasculares e diabetes. Para isso nos utilizamos de uma técnica estatística denominada regressão logística multinomial, que permite a análise da relação das variáveis explicativas com a variável resposta em seus vários níveis. O estudo identificou a associação de variáveis relativas aos hábitos alimentares, o percentual e a distribuição de gordura à obesidade e ao risco para diabetes e doença cardiovascular.

Palavras-chave: *Obesidade; Risco Cardiovascular; Regressão Logística; Modelos multinomiais.*

1. Introdução

A obesidade vem se apresentando como um problema de saúde de âmbito mundial, abrangendo jovens e adultos, homens e mulheres de todas as classes sociais. A Organização Mundial da Saúde – OMS define a obesidade como uma doença caracterizada pelo acúmulo

¹luciana.nucci@terra.com.br

²anacarol.nunes@gmail.com

³liaokawa@gmail.com

⁴ricacordeiro@gmail.com

⁵luyoka@hotmail.com

de gordura corporal em excesso, que traz consequências à saúde das pessoas (WHO. World Health Organization, 1998). Zanardi define obesidade como uma doença crônica consequente ao excesso de gordura armazenado sob forma de triglicérides, resultado da ingestão energética aumentada em relação ao gasto dessa gordura (Zanardi, 2006). Estudos revelam que ela atinge em torno de 26% da população maior de 18 anos em países desenvolvidos, como os Estados Unidos, variando de 18 a 31% em cidades latino-americanas, chegando a atingir em torno de 10% no Brasil (CDC. Centers for Disease Control and Prevention, 2007; Cuevas et al., 2009; IBGE. Instituto Brasileiro de Geografia e Estatística, 2004). Devido à estreita relação do excesso de peso com a ocorrência de agravos crônicos não transmissíveis e ao grau epidêmico de sua expansão, a obesidade é considerada a mais importante desordem nutricional da atualidade, estando entre os problemas de saúde pública de maior importância mundial (WHO. World Health Organization, 1998).

Sabe-se que hábitos alimentares saudáveis aliados às atividades físicas podem prevenir a condição do indivíduo obeso, sendo essa a melhor forma de controle da doença, pois apesar de extensa literatura acerca dos diferentes tratamentos para a obesidade, que incluem dieta, atividade física, medicamentos, mudança comportamental e até a cirurgia bariátrica, ainda não foi possível alcançar uma conclusão definitiva (Cowburn et al., 1997; Francischi et al., 2000).

Por trazer, associado a ela, problemas cardiovasculares, articulares e até mesmo sociais, a obesidade tem proporcionado muitos estudos com o intuito de desvendar suas relações causais que é atravessado pela necessidade de encontrar respostas para além da identificação de fatores de risco biológicos, de modo a inserir esse fenômeno no contexto socioambiental em que tem lugar, para o que se faz necessária a utilização de ferramental disponível em diferentes áreas do conhecimento (Huang e Glass, 2008).

Em estudos epidemiológicos muito frequentemente a relação entre exposição e adoecimento é modelada pela distribuição binomial, por meio de ajustes de modelos de regressão linear generalizada, tendo como função de ligação o logito da probabilidade de doença. Entretanto, nem sempre é conveniente classificar dicotomicamente a população estudada. Pode ser interessante, para os propósitos do estudo, subdividir os casos analisados de acordo com características tais como gravidade, tipo celular ou forma clínica, obtendo-se maior compreensão do desfecho e sua relação com as exposições estudadas. Nestas situações, modelos que utilizam a distribuição binomial não são os melhores métodos a serem utilizados. Uma opção natural para a análise são modelos lineares generalizados com resposta multinomial, uma extensão da distribuição binomial para mais de dois resultados possíveis (Agresti, 2002; Ananth e Kleinbaum, 1997). Portanto, propõe-se nesse estudo a utilização da técnica de regressão logística multinomial para a identificação dos fatores de risco para obesidade (Mafra et al., 2010).

2. Objetivos

Diante deste panorama mundial, o estudo a seguir procurou identificar alguns fatores de risco para obesidade, classificada em mais de duas categorias, associados direta ou indiretamente a essa doença que tem se tornado cada vez mais preocupante na sociedade.

3. Metodologia

3.1. Coleta dos Dados

Os dados usados provêm de um estudo epidemiológico transversal, com dados coletados entre outubro de 2006 e dezembro de 2007 (Zangirolani, 2010). A população fonte foi composta por moradores do Distrito Sul de Campinas-SP, com idades entre 19 e 65 anos. Campinas é considerada polo industrial e tecnológico do estado de São Paulo e o terceiro polo industrial do Brasil, localizada a Noroeste da capital paulista, a aproximadamente 100 km de distância, com uma população estimada de mais de um milhão de habitantes em 2007. O Distrito Sul, local onde foi realizado o estudo, abrange uma área de 127,8 km² e possui uma população estimada de 277.400 habitantes (IBGE. Instituto Brasileiro de Geografia e Estatística, 2007).

A amostra foi obtida por procedimentos aleatórios em dois estágios. No primeiro, foram sorteados os domicílios a partir do cadastro feito por Agentes Comunitários de Saúde. No segundo, foram catalogados os moradores dos domicílios visitados e, dentre estes, foi sorteado, um morador em cada residência para participar do estudo, mediante assinatura de um termo de consentimento livre e esclarecido. Foram visitados 730 domicílios, compondo uma amostra de 649 adultos estudados.

Os dados foram coletados por meio de entrevistas domiciliares, utilizando-se um questionário no qual foram registrados dados de identificação, medidas antropométricas e sociodemográficas. Para a coleta dos dados antropométricos foram seguidos os métodos propostos pelo *Anthropometric standardization reference manual* (Lohman et al., 1991), utilizando uma balança digital portátil com capacidade de 150 kg e precisão de 100g, e a estatura foi aferida com estadiômetro portátil, devidamente afixado em parede ou porta, sem rodapé, aliado a um esquadro de ângulo reto, utilizado como régua no topo da cabeça do indivíduo para dar precisão à medida.

3.2. Variável Resposta

Para os ajustes logísticos, optou-se pela análise de dois indicadores como variável resposta, são eles:

1. Índice de Massa Corporal (IMC), obtido através da razão $\text{Peso}(\text{kg})/(\text{Altura}(\text{m}))^2$, categorizado como eutrofia ($\text{IMC} < 25\text{kg}/\text{m}^2$), sobrepeso ($25\text{kg}/\text{m}^2 \leq \text{IMC} < 30\text{kg}/\text{m}^2$) e obesidade ($\text{IMC} \geq 30\text{kg}/\text{m}^2$);
2. Risco para diabetes e doença cardiovascular (RCV), segundo recomendação da Associação Brasileira para o Estudo da Obesidade e da Síndrome Metabólica (ABESO. Associação brasileira para o estudo da obesidade e da síndrome metabólica, 2009), categorizado de acordo com os intervalos de IMC e Circunferência abdominal apresentados na Tabela 1.

Tabela 1: Combinação das medidas de circunferência abdominal e IMC para avaliar obesidade e risco para diabetes 2 e doença cardiovascular

IMC (kg/m^2)	Circunferência abdominal (cm)			
	Homem		Mulher	
	94 – 102	> 102	80 – 88	> 88
< 25	Normal	Aumentado	Normal	Aumentado
≥ 25 a < 30	Aumentado	Alto	Aumentado	Alto
≥ 30	Alto	Muito alto	Alto	Muito Alto

Fonte: Associação Brasileira para o estudo da obesidade e da síndrome metabólica (ABESO), 2009.

3.3. Análise Estatística

Foram criados dois modelos multinomiais simples relacionados a cada uma das variáveis relativa à obesidade. Em cada modelo, foram testadas todas as variáveis preditoras independentes através do teste-t bilateral, que, devido ao tamanho da amostra pode ser aproximado pela distribuição Normal, adotando-se um nível de significância de 20%. Este nível foi escolhido com o intuito de não haver uma rejeição muito grande das variáveis, já que nesta etapa queremos apenas descartar as variáveis que tem pouquíssima relação com os desfechos. Em seguida, foram feitos os modelos múltiplos.

Para a obtenção dos modelos finais utilizamos o método *backward* como método de seleção de variáveis regressoras ao nível de significância de 5% em pelo menos um dos log-itos estimados, iniciando-se com um modelo onde constavam variáveis significativas a 20% no modelo simples, independentes e clinicamente relevantes. Deste, foram retiradas, uma a uma, aquelas que apresentassem os maiores p-valores.

As análises foram realizadas no pacote estatístico R versão 2.11.1.

4. Resultados

A amostra apresentou uma predominância do sexo feminino, representado por 55,78% dos indivíduos pesquisados, média de idade de 41,6(\pm 12,5) anos e escolaridade média de 8,6(\pm 4,2) anos de estudo. A frequência de critérios de risco relativos à obesidade é apresentada na Tabela 2.

Tabela 2: Frequência da classificação dos indivíduos segundo critérios de risco relativos à obesidade. Distrito Sul de Campinas, 2007

Classificação	Frequência	%
Eutrofia (IMC < 25)	300	46
Sobrepeso ($25 \leq$ IMC < 30)	212	33
Obesidade (IMC \geq 30)	137	21
Risco Cardiovascular		
Normal	287	44
Aumentado	157	24
Alto	96	15
Muito alto	109	17

No modelo 1, onde avaliamos quais variáveis preditoras estão relacionadas à variável resposta IMC categorizada em eutrofia, sobrepeso ou obesidade, obtivemos que as variáveis idade, esforço no trabalho, número de refeições ao dia, circunferência da cintura, circunferência abdominal média, circunferência do quadril, dobra tricipital, dobra bicipital, dobra subescapular, dobra supra ilíaca, razão entre as medidas da cintura e as do quadril (RCQ) e porcentagem de gordura foram significativas na análise simples. Destas, não foram consideradas para o modelo múltiplo a circunferência da cintura e do quadril por serem, conjuntamente, colineares com a RCQ, nem as dobras, pois também são, conjuntamente, colineares com a porcentagem de gordura.

No modelo 2, onde avaliamos quais variáveis preditoras estão relacionadas à variável resposta risco para diabetes e doença cardiovascular classificado em normal, aumentado, alto e muito alto, obtivemos que as variáveis idade, esforço no trabalho, circunferência do quadril, dobra tricipital, dobra bicipital, dobra subescapular, dobra supra ilíaca, RCQ e porcentagem de gordura foram significativas. Como no modelo 1, as variáveis colineares não foram consideradas para o modelo múltiplo.

Notamos que há uma reincidência de variáveis nos dois modelos considerados. Isso nos indica que dentre todas as variáveis preditoras analisadas no começo do projeto essas, que foram consideradas importantes em pelo menos um dos modelos, podem ser relevantes na identificação dos fatores de risco para a obesidade no Distrito Sul de Campinas. Assim,

a partir delas foram obtidos os modelos multinomiais múltiplos apresentados nas Tabelas 3 e 4.

A Tabela 3 mostra a relação entre as variáveis que permaneceram no modelo final com IMC. Nota-se que a região Nordeste está associada a um maior risco de sobrepeso e variáveis tradicionalmente indicativas de obesidade, como circunferência abdominal e percentual de gordura, também apresentaram uma associação significativa. O percentual de gordura mostrou interação com a variável *ceia*, que indica se o indivíduo realiza refeição noturna, apresentando associação de maior risco para quem não tem esse hábito e possui percentual de gordura mais elevado, tanto para sobrepeso, quanto para obesidade.

Tabela 3: Modelo de regressão logística multinomial da associação entre categorias de IMC e variáveis explicativas associadas. Distrito Sul de Campinas, SP, 2007

	<i>Sobrepeso</i>		<i>Obeso</i>	
	Razão de Chances	IC 95%¹	Razão de Chances	IC 95%¹
Região Nordeste ²	3,825	1,426 – 10,265	1,743	0,911 – 3,335
Região Noroeste ²	1,348	0,495 – 3,667	1,329	0,683 – 2,588
Região Sul ²	2,394	0,643 – 8,912	1,039	0,419 – 2,579
Circunf. Abdominal (cm)	1,384	1,311 - 1,461	1,200	1,156 - 1,246
% Gordura (não ceia)	1,962	1,279 - 3,010	1,463	1,218 - 1,758
% Gordura (ceia)	1,633	1,091 - 2,445	1,115	0,956 - 1,300

¹ Intervalo de Confiança de 95%.

² Comparadas à Região Centro.

³ Todas as categorias de resposta comparadas com *Eutrófico*.

Do mesmo modo, a Tabela 4 mostra as estimativas obtidas pelo modelo em que a resposta é o RCV. Nesse modelo a idade, o percentual de gordura e a RCQ estão associados a maiores riscos e as variáveis relacionadas à alimentação, número de refeições diárias e o consumo de bolachas, chicletes e chocolate estão associados a menores riscos.

4.1. Análise dos Modelos Multinomiais Múltiplos

Nos modelos de regressão linear, as medidas resumos de ajuste são funções do resíduo definido como a diferença entre o valor observado e o valor ajustado. Nos modelos de regressão logística existem muitas maneiras de se medir essa diferença. Uma dessas medidas de diferença entre o valor observado e o valor ajustado é o desvio residual (Hosmer e Lemeshow, 2000). Podemos realizar um teste para avaliar a qualidade de ajuste,

Tabela 4: Modelo de regressão logística multinomial da associação entre categorias de Risco Cardiovascular e variáveis explicativas associadas. Distrito Sul de Campinas, SP, 2007

	<i>Aumentado</i>		<i>Alto</i>		<i>Muito Alto</i>	
	RC ¹	IC 95% ²	RC ¹	IC 95% ²	RC ¹	IC 95% ²
Idade (anos)	1,074	1,037 – 1,111	1,07	1,041 – 1,101	1,026	1,006 – 1,047
Refeições/dia	0,642	0,426 – 0,969	0,596	0,421 – 0,843	0,813	0,633 – 1,045
Doces						
Não vs Sim	0,567	0,195 – 1,645	0,64	0,261 – 1,572	0,597	0,318 – 1,122
Às Vezes vs Sim	0,384	0,138 – 1,068	0,484	0,208 – 1,128	0,363	0,199 – 0,659
%Gordura	3,33	2,752 – 4,03	1,926	1,679 – 2,209	1,378	1,273 – 1,491
RCQ ³						
Moderado vs Baixo	2,236	0,65 – 7,691	3,407	1,327 – 8,746	1,578	0,903 – 2,759
Alto vs Baixo	3,859	1,142 – 13,039	7,141	2,783 – 18,325	1,617	0,886 – 2,948

¹ Razão de Chances comparada à categoria Normal.

² Intervalo de Confiança de 95%.

³ Razão Cintura/Quadril.

partindo de uma estatística baseada no desvio residual. A análise deste teste é baseada na comparação do desvio residual do modelo múltiplo completo com a do modelo apenas com o intercepto e segue uma distribuição Qui-Quadrado com $n_0 - n_c$ graus de liberdade (g.l.), onde n_0 são os g.l. residuais do modelo nulo e n_c os g.l. residuais do modelo completo. A hipótese nula deste teste assume que todos os parâmetros $\beta = 0$ e a alternativa é que pelo menos um deles difere de zero.

Os valores obtidos para essas estatísticas estão descritos na Tabela 5.

Tabela 5: Análise de Deviance dos modelos propostos

Modelo	Deviance	G.L.	Deviance	G.L.	P-valor
	Residual	Residual			
1	692,65	1262	645,24	14	<0,0001
nulo 1	1337,9	1276			
2	1055,33	1878	562,92	21	<0,0001
nulo 2	1618,25	1899			

Assim, podemos concluir que os ajustes dos modelos são satisfatórios devido aos baixos p-valores, o que indica que rejeitamos a hipótese de que todos os betas são iguais

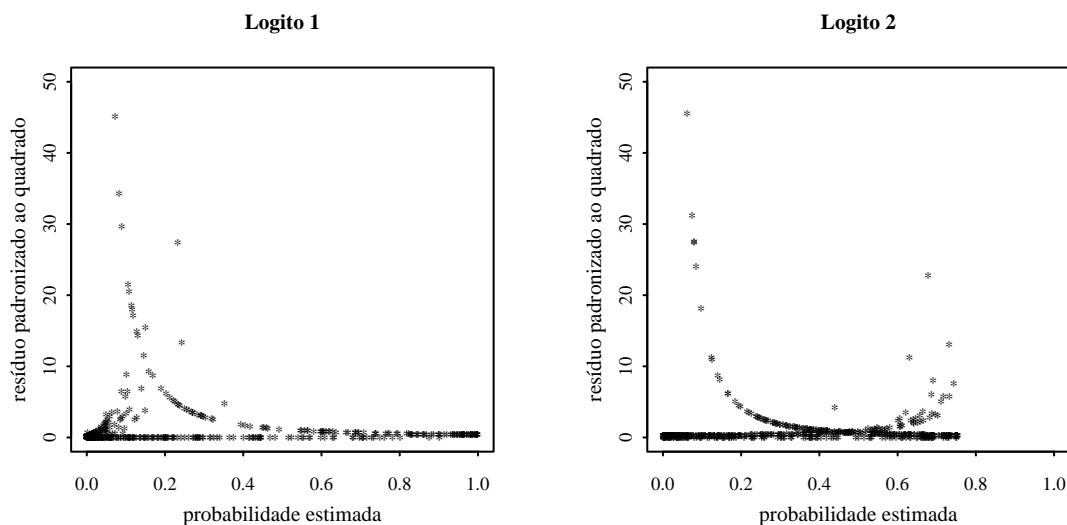


Figura 1: Dispersão dos resíduos padronizados ao quadrado em função das probabilidades estimadas pelo Modelo 1 (variável resposta IMC).

a zero.

Para melhor verificar a qualidade dos ajustes dos modelos, uma análise gráfica dos resíduos também foi realizada. As Figuras 1 e 2 mostram os gráficos das probabilidades estimadas pelos valores de resíduos padronizados elevados ao quadrado para os Modelos 1 e 2, que tem como variável resposta o IMC e o RCV, respectivamente. Analisando as Figuras, vemos que os modelos estão bem ajustados, pois apresentam poucos pontos com valores de resíduo elevados.

Para conferir os pressupostos sobre a normalidade dos resíduos foram realizados testes de normalidade de Shapiro-Wilk, os quais deram altamente significativos, validando assim os modelos propostos (Royston, 1982).

5. Conclusões

O modelo multinomial aplicado a dados epidemiológicos mostra-se como importante ferramenta de análise, pois foi possível identificar fatores associados ao sobrepeso que não necessariamente estão associados à obesidade, como a localização na região estudada e o percentual de gordura interagindo com refeição noturna.

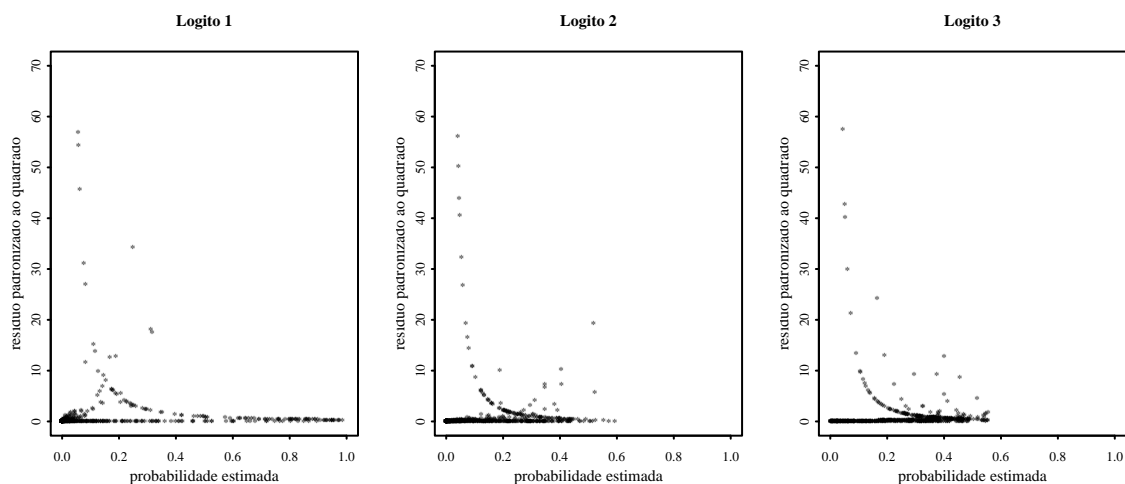


Figura 2: Dispersão dos resíduos padronizados ao quadrado em função das probabilidades estimadas pelo Modelo 2 (variável resposta RCV).

Ao avaliarmos a diabetes e doença cardiovascular (RCV) identificamos variáveis tradicionalmente associadas aos desfechos, como a idade, o número de refeições por dia, o consumo de doces, o percentual de gordura corporal e a razão cintura–quadril. O fracionamento das refeições ao longo do dia foi identificado como fator protetor para RCV aumentado e alto, enquanto o consumo de doces moderado foi associado apenas ao RCV muito alto. A razão cintura–quadril também esteve associada apenas ao RCV nas categorias moderado e alto.

Os modelos propostos neste trabalho nos permitiram concluir que a obesidade e os problemas de saúde associados a ela podem estar relacionados aos hábitos alimentares dos indivíduos que resultam em acúmulo de gordura corporal/abdominal, elevando o risco de agravos à saúde decorrentes dessa condição. A identificação da Unidade de Saúde associada ao sobrepeso aponta para diferenças na distribuição do IMC na área estudada. No estudo foi possível identificar que a região Nordeste apresentou associação de maior risco de sobrepeso. Uma investigação de quais características da região, como por exemplo a condição social, infra–estrutura e qualidade de vida, deve ser feita com o intuito de propor políticas para minimizar o risco de obesidade e saúde.

Agradecimentos

Agradecemos à Fundação de Amparo à Pesquisa do Estado de São Paulo - FAPESP (processos: 2006/05920-7; 2007/07435-1; 2008/58461-5).

Referências

- ABESO. Associação brasileira para o estudo da obesidade e da síndrome metabólica (2009). *Diretrizes brasileiras de obesidade 2009/2010*. 3. ed. AC Farmacêutica, Itapevi, SP.
- Agresti, A. (2002). *Categorical data analysis*. 2. ed. John Wiley & Sons, New York.
- Ananth, C. e Kleinbaum, D. (1997). Regression models for ordinal responses: a review of methods and applications. *International Journal of Epidemiology*, 26:1323–1333.
- CDC. Centers for Disease Control and Prevention (2007). Behavioral risk factor surveillance system survey. Morbidity and Mortality Weekly Report. Disponível no site: www.cdc.gov/mmwr/preview/mmwrhtml/mm5728a1.htm.
- Cowburn, G., Hillsdon, M., e Hankey, C. (1997). Obesity management by life-style strategies. *British Medical Bulletin*, 53:389–408.
- Cuevas, A., Alvarez, V., e Olivos, C. (2009). The emerging obesity problem in latin américa. *Expert Review of Cardiovascular Therapy*, 7:281–288.
- Francischi, R., Pereira, L., Freitas, C., Klopfer, M., Santos, R., Vieira, P., e Lancha Júnior, A. (2000). Obesidade: atualização sobre sua etiologia, morbidade e tratamento. *Revista de Nutrição*, 13:17–28.
- Hosmer, D. e Lemeshow, S. (2000). *Applied Logistic Regression*. John Wiley & Sons, Canadá.
- Huang, T. e Glass, T. (2008). Transforming research strategies for understanding and preventing obesity. *Journal of the American Medical Association*, 300:1811–1813.
- IBGE. Instituto Brasileiro de Geografia e Estatística (2004). Pesquisa de orçamentos familiares 2002-2003: análise da disponibilidade domiciliar de alimentos e do estado nutricional no brasil. Rio de Janeiro.
- IBGE. Instituto Brasileiro de Geografia e Estatística (2007). População recenseada e estimada, segundo municípios. São Paulo. Disponível no site: www.ibge.gov.br/home/estatistica/populacao/contagem2007.

- Lohman, T., Roche, A., e Martorell, R. (1991). *Anthropometric standardization reference manual*. Human Kinetics Books, Champaign.
- Mafra, A., Nucci, L., Cordeiro, R., e Stephan, C. (2010). Estudo caso-controle com resposta multinomial: uma proposta de análise. *Cadernos de Saúde Pública*, 26:451–459.
- Royston, P. (1982). An extension of shapiro and wilk's w test for normality to large samples. *Applied Statistics*, 31:115–124.
- WHO. World Health Organization (1998). Obesity: preventing and managing the global epidemic. Report of WHO consultation on obesity, Geneva.
- Zanardi, G. (2006). Obesidade infantil: Prevalência e atividade física. Monografia, UNICAMP, Campinas.
- Zangirolani, L. (2010). *Topologia do excesso de peso no Distrito Sul de Campinas, São Paulo*. Tese de Doutorado, UNICAMP.

