

## Obesidade adulta nas nações: uma análise via modelos de regressão beta

Adult obesity in different countries: an analysis via beta regression models

Obesidad adulta en naciones: un análisis mediante modelos de regresión beta

Saul de Azevêdo Souza <sup>1</sup>  
Allan Batista Silva <sup>1</sup>  
Ulanna Maria Bastos Cavalcante <sup>1</sup>  
Caliandra Maria Bezerra Luna Lima <sup>1</sup>  
Tatiene Correia de Souza <sup>1</sup>

doi: 10.1590/0102-311X00161417

### Resumo

A obesidade é considerada um grave problema de saúde pública, por se tratar de uma doença epidêmica de grande repercussão no cenário mundial e que está relacionada ao desenvolvimento de outras doenças crônicas, como, por exemplo, hipertensão, diabetes e doenças cardiovasculares. Diante disso, o presente trabalho tem como objetivo estudar a distribuição da obesidade em adultos de diferentes nações, por meio do modelo de regressão beta. Trata-se de um estudo ecológico descritivo com abordagem quantitativa e inferencial com foco na análise de regressão beta. A aplicação desse método considerou um conjunto de dados reais, obtidos a partir de fontes de informação pública, referente à obesidade adulta nas nações no ano de 2014. Após a análise descritiva dos dados, verificou-se que 50% das nações apresentam uma prevalência de adultos obesos maiores que 0,20 (20%). Adicionalmente, ao analisar a distribuição de sua prevalência por nação, constatou-se que os menores valores de obesidade adulta estão concentrados nos países pertencentes aos continentes da Ásia e África. Por outro lado, os maiores valores encontram-se distribuídos entre os países nos continentes da América e Europa. Ainda, a partir da análise gráfica do boxplot, foram observadas evidências de uma possível diferença nas proporções de adultos obesos entre os continentes da América e Europa com os da África e Ásia. Após ajustar o modelo de regressão beta com dispersão variável, foi possível identificar, ao nível de 5% de significância, que as variáveis consumo médio de álcool em litros por pessoa, porcentagem de atividade física insuficiente, porcentagem da população que vive em áreas urbanas e expectativa de vida apresentam efeito.

Obesidade; Doença Crônica; Modelos Lineares

### Correspondência

S. A. Souza  
Universidade Federal da Paraíba.  
Cidade Universitária s/n, João Pessoa, PB 58051-085, Brasil.  
saul\_asouza@hotmail.com

<sup>1</sup> Universidade Federal da Paraíba, João Pessoa, Brasil.



## Introdução

### Obesidade adulta no cenário mundial

A obesidade é considerada uma doença epidêmica de grande repercussão no cenário mundial, recorrente tanto em países desenvolvidos como naqueles em desenvolvimento <sup>1,2</sup>. Tal doença pode apresentar como causas os fatores genéticos, metabólicos, ambientais, sociais, culturais, econômicos, estilo de vida ou ainda estar relacionada a fatores demográficos <sup>3,4</sup>.

O índice de massa corporal (IMC) é uma medida capaz de avaliar a concentração de gordura em adultos, sendo definido como a razão entre o peso do indivíduo dado em quilogramas (kg) e sua altura ao quadrado (m<sup>2</sup>) <sup>5</sup>. Assim, as pessoas que apresentam  $IMC \geq 30 \text{ kg/m}^2$  são classificadas como obesas.

A Organização Mundial da Saúde (OMS) define a obesidade como a excessiva concentração de gordura que pode prejudicar a saúde do indivíduo <sup>5</sup>. Portanto, o consumo de alimentos altamente energéticos e a falta de atividade física se destacam por facilitarem o ganho de calorias e diminuir o gasto de energia corporal ao longo do dia, tornando a balança energética do indivíduo positiva e facilitando o acúmulo de gordura <sup>6</sup>.

Além de inserida no grupo de doenças crônicas não transmissíveis (DCNT), a obesidade é considerada um dos importantes fatores de risco para outras complicações como: diabetes mellitus, hipertensão, doenças cardiovasculares etc. <sup>7,8</sup>. As DCNT, em especial as citadas anteriormente, representam um sério problema de saúde pública, pois são apontadas como sendo as principais causas de mortes no mundo <sup>9</sup>. Em 2008, por exemplo, as DCNTs foram responsáveis por cerca de 63% das mortes no mundo, sendo 80% delas ocorridas em países de baixa e média renda <sup>10</sup>.

A obesidade é uma doença de grande impacto social, familiar e financeiro principalmente para as famílias das pessoas acometidas. Os tratamentos oferecidos aos obesos – a fim de cuidar das consequências decorrentes da doença – representam enormes gastos no setor da saúde. No Brasil, por exemplo, os custos totais com os procedimentos associados ao sobrepeso e à obesidade custam cerca de 2,1 bilhões de dólares anualmente <sup>11</sup>. Já os Estados Unidos são um dos países que mais sofrem com os problemas decorridos da obesidade, pois cerca de 35% de sua população é considerada obesa, e seus gastos para tratar dessa doença ultrapassam bilhões de dólares anualmente <sup>12</sup>.

A Organização para a Cooperação e Desenvolvimento Económico (OECD) é uma organização internacional constituída por 34 países – tanto desenvolvidos como em desenvolvimento – que tem como objetivo promover políticas que melhorem a economia e o bem-estar social das pessoas ao redor do mundo. Essa organização, em seu relatório para o ano de 2014, mostrou que, nos últimos cinco anos, os países Canadá, Inglaterra, Itália, Coreia do Sul, Espanha e Estados Unidos apresentaram um crescimento modesto ou praticamente estável do sobrepeso e da obesidade. Por outro lado, os países Austrália, França, México e Suíça apresentaram um crescimento de 2% a 3%, não havendo nenhum indício da redução ou contenção dessa epidemia entre as nações. Além disso, é estimado que os gastos das nações no setor de saúde relacionados a essa doença variam de 1% a 3%, podendo ser maiores quando associados a outras complicações <sup>13</sup>.

Portanto, por se tratar de um problema mundial que envolve diversos países, inclusive o Brasil, é necessário compreender um pouco mais sobre a distribuição da obesidade em nível mundial e identificar os possíveis fatores relacionados ao seu aumento nos últimos anos. Para tanto, muitos autores recorrem aos modelos de regressão logística, que têm apresentado bastante destaque nos estudos epidemiológicos, a fim de identificar alguma relação entre as variáveis envolvidas, em um contexto que a variável resposta é dicotômica e a unidade de interesse são os indivíduos <sup>14,15</sup>. No entanto, o presente estudo tem como objetivo estudar a distribuição da obesidade em adultos de diferentes nações, por meio do modelo de regressão beta. Essa abordagem é válida uma vez que a variável resposta se trata de uma proporção definida no intervalo (0,1).

### Modelos clássicos de regressão e o modelo de regressão beta

Na literatura, existe uma ampla quantidade de métodos estatísticos que podem ser utilizados para modelar dados. Contudo, na maioria das vezes, o que se vê é o uso indiscriminado do modelo de regressão logística. Para tanto, é útil conhecer os diferentes tipos de modelos propostos na literatura

para que a análise da relação entre a variável resposta e as demais covariáveis independentes seja feita da melhor maneira possível.

Em diversas situações práticas, sejam elas observacionais ou experimentais, pesquisadores buscam compreender e explicar os fenômenos ocorridos em diversas áreas da ciência. Para isso, é possível utilizar os modelos de regressão, pois esses permitem expressar a relação existente entre uma variável resposta,  $Y_t$ , e as  $p$  covariáveis independentes,  $(X_{1t}, \dots, X_{pt})$ , abordadas no estudo. O modelo de regressão linear é um dos métodos mais conhecidos, devido à facilidade de interpretação dos seus parâmetros por parte dos pesquisadores, além de se encontrar disponível em diversos programas estatísticos. A estrutura de regressão desse modelo pode ser definida da seguinte maneira:

$$Y_t = \beta_0 + \beta_1 X_{1t} + \dots + \beta_p X_{pt} + \varepsilon_t$$

com  $t = 1, \dots, n$ , em que  $n$  é o total de observações no estudo. Aqui,  $Y_t$  é a variável resposta ou desfecho,  $(X_{1t}, \dots, X_{pt})$  são as covariáveis independentes, e  $(\beta_0, \dots, \beta_p)$  são os parâmetros desconhecidos a serem estimados. Tem-se que os erros,  $\varepsilon_t$ , são uma sequência aleatória, independente e normalmente distribuída com média zero e variância constante. De forma sintética, os modelos de regressão buscam descrever a relação existente entre variáveis por meio de uma equação matemática <sup>16</sup>.

Kieschnick & McCullough <sup>17</sup>, estudando a modelagem de variáveis restritas ao intervalo (0,1), foram capazes de identificar sete tipos de modelos utilizados na literatura para analisar dados com restrição no intervalo (0,1). São eles o modelo normal linear, o modelo *logito*, o modelo normal censurado, o modelo normal não linear, o modelo baseado na distribuição beta, o modelo baseado na distribuição simplex e o modelo de quasi-verossimilhança. Os autores ainda discutem o uso inapropriado do estimador de mínimos quadrados ordinários nesse cenário. Por fim, eles recomendam o uso de regressão baseada na distribuição beta ou um modelo de regressão quasi-verossimilhança <sup>18</sup> para dados com esse tipo de restrição.

Dessa forma, para modelar dados assimétricos e restritos ao intervalo (0,1), Ferrari & Cribari-Neto <sup>19</sup> propuseram o modelo de regressão beta. Essa classe de modelos assume que a distribuição de probabilidade da variável resposta é a beta, ou seja, os dados devem estar dispostos como taxas ou proporções, equivalentes às prevalências em estudos epidemiológicos. Diferente dos modelos normais lineares, seu estimador usual é o de máxima verossimilhança. Dessa forma, é possível estimar o vetor de parâmetros desconhecidos com base na função de verossimilhança. O modelo linear normal não pode ser utilizado quando os dados contêm valores zeros e/ou uns, ou seja, quando alguma observação equivale aos limites do intervalo. Isso se dá pelo fato de que proporções no intervalo (0,1) não estão definidas sobre todos os números reais, que é um dos pressupostos para a distribuição normal – principal característica assumida pela variável para que seja possível aplicar o modelo linear <sup>20</sup>.

Nesse cenário, a função de *log*-verossimilhança do modelo de regressão beta se torna ilimitada. Além disso, não é adequado assumir que os dados são provenientes de uma distribuição absolutamente contínua. Para tanto, uma solução adequada seria utilizar o modelo de regressão beta inflacionado em zero ou um, em que a distribuição da variável resposta é uma mistura entre uma distribuição Bernoulli e uma distribuição beta <sup>20</sup>.

Na estrutura de regressão para modelar a resposta média, temos que a resposta média  $y_t$  está relacionada a um preditor linear,  $\eta_t$ , por meio de uma função de ligação da seguinte maneira:

$$g(\mu_t) = \sum_{i=1}^k X_{ti} \beta_i = \eta_t$$

em que  $\beta = (\beta_1, \dots, \beta_k)^T$  é o vetor de parâmetros desconhecidos a ser estimado,  $X = (X_{1t}, \dots, X_{kt})$  são observações de  $k$  variáveis independentes. Aqui, a resposta média é obtida aplicando a inversão da função de ligação  $g(\cdot)$ , ou seja,  $\mu_t = g^{-1}(\eta_t)$ .

Vale ressaltar que esse modelo considera o parâmetro de precisão constante ao longo das observações. Contudo, em certas situações, esse parâmetro pode variar ao longo das observações <sup>21,22,23,24,25</sup>. Ou seja, o parâmetro de precisão é variável e precisa ser modelado a partir de uma estrutura de regressão similar ao da resposta média. A estrutura de regressão da precisão passa a ser definida como:

$$h(\phi_t) = \sum_{j=1}^q Z_{tj} \gamma_j = \vartheta_t$$

em que  $\gamma = (\gamma_1, \dots, \gamma_q)^T$  é um vetor de parâmetros desconhecidos,  $Z = (Z_{11}, \dots, Z_{1q})$  são observações de  $q$  variáveis independentes ( $k + q < n$ ),  $\vartheta_i$  é o preditor linear, e  $h(\cdot)$ , uma função de ligação. Existem algumas escolhas possíveis para as funções de ligação  $g(\cdot)$  e  $h(\cdot)$ . Por exemplo, para  $g(\cdot)$ , referente ao modelo da média, pode-se utilizar a função de ligação logit,  $g(\mu) = \log\left(\frac{\mu}{1-\mu}\right)$  ou cloglog,  $g(\mu) = \log\{-\log(1-\mu)\}$ . Em relação ao modelo da precisão, pode-se utilizar a função  $h(\varnothing) = \log(\varnothing)$  ou  $h(\varnothing) = \sqrt{\varnothing}$  para  $2^6$ .

O conceito de heteroscedasticidade, não constância da variância dos erros, aplicado ao modelo de regressão beta difere do empregado ao modelo normal, que utiliza frequentemente a variância como medida de dispersão. De fato, mesmo que o parâmetro de dispersão seja constante, a variância da variável resposta não é constante, pois depende das médias desconhecidas que variam conforme a estrutura do modelo. Naturalmente, a dispersão é tratada como o inverso da precisão, ou seja, quanto maior a dispersão dos dados ao longo das observações menor a precisão da resposta média, e vice versa. Além disso, a modelagem correta da dispersão tem influência direta sobre os parâmetros da estrutura da média, o que melhora os resultados inferenciais.

## Metodologia

Trata-se de um estudo ecológico descritivo, com abordagem quantitativa e inferencial com foco na análise de regressão. Os dados utilizados são referentes à obesidade adulta em 78 países no ano de 2014, em que, para o cálculo da proporção observada, foi considerada a população adulta, com 18 anos ou mais, que apresentasse IMC  $> 30\text{kg/m}^2$ . A amostra foi constituída por 78 observações (proporções) referentes às nações ao redor do mundo, das quais, 25 (32%) pertencem à África, 11 (14%) pertencem à América, 14 (18%) pertencem à Ásia, 25 (32%) pertencem à Europa, e 3 (4%) pertencem à Oceania.

Os dados foram coletados nas bases eletrônicas do Banco Mundial (<http://databank.worldbank.org>) e OMS (<http://www.who.int>). Essa primeira base é composta por cinco instituições que buscam reduzir a pobreza e fornecer assistência técnica e financeira aos países em desenvolvimento. A segunda base consiste numa organização que funciona em mais de 150 países e conta com os governos e outros parceiros para garantir o mais alto nível possível de saúde para as pessoas.

Os dados coletados foram tabulados em planilha eletrônica e submetidos ao software R (The R Foundation for Statistical Computing; <http://www.r-project.org>). Esse software é uma plataforma livre que possui diversos métodos estatísticos de análise de dados já implementados. Vale ressaltar que foram coletados os dados mais atualizados que estavam disponíveis e que abrangiam um maior número de nações. Além do mais, por se tratar de um banco de dados de domínio público, não foi necessário submeter o projeto ao Comitê de Ética em Pesquisa.

Inicialmente, foi feita uma análise descritiva dos dados a fim de extrair informações importantes a respeito das variáveis abordadas no estudo. As variáveis citadas, nesta pesquisa, são apresentadas abaixo com suas respectivas descrições:

*OB2014*: proporção de adultos obesos, 18 anos ou mais, com IMC =  $30\text{kg/m}^2$  em 2014;

*INAT*: porcentagem de atividade física insuficiente entre os adultos em 2010. Em outros termos, porcentagem de população definida atingindo menos de 150 minutos de atividade física de intensidade moderada por semana ou menos de 75 minutos de atividade física de intensidade vigorosa por semana, ou equivalente;

*EDUC*: gastos com a educação como porcentagem da despesa total do governo em 2010;

*VIDA*: expectativa de vida ao nascer em anos no ano de 2014;

*ALC*: média do consumo em litros de álcool puro por pessoa em um ano, considerando a população com 15 anos ou mais em 2008;

*URB*: porcentagem da população que vive em áreas urbanas em 2014.

Em seguida, realizaram-se os procedimentos inferenciais e as medidas de qualidade do ajuste relacionado ao modelo de regressão beta, em que, para tal modelo, utilizou-se o pacote *betareg* do programa estatístico utilizado. Conforme apresentado anteriormente, o modelo de regressão beta com dispersão variável apresenta, como diferencial, a possível modelagem da variabilidade dos dados, artifício que permite melhorar os resultados inferenciais. Além disso, esse modelo foi escolhido pois as variáveis estudadas são fornecidas em proporção. Somando-se a isso, o modelo de regressão beta

tem a vantagem de possibilitar a expansão das conclusões a respeito do tema em estudo estimando o impacto de uma determinada covariável sobre a resposta média.

## Resultados e discussão

A análise descritiva dos dados pode ser observada na Tabela 1, na qual é apresentado o valor mínimo, primeiro quartil ( $Q_{1/4}$ ), mediana, média, terceiro quartil ( $Q_{3/4}$ ), máximo e coeficiente de variação (CV) das variáveis utilizadas na modelagem de regressão beta. Diante dessa tabela, podemos observar que a proporção de adultos obesos varia de 0,03 até 0,41, com cerca de 25% dessas nações apresentando valores de *OB2014* superior a 0,26 ou 26%.

Em 50% das nações, a prevalência das pessoas que praticam atividade física insuficiente é superior a 23,8%, com mínimo de 4,1% e máximo de 63,6%. A menor expectativa de vida ao nascer foi de 49 anos, e a maior, de 83 anos, com uma esperança de viver, em média, 72 anos. Os gastos com a educação como porcentagem da despesa total pelo governo variou de 5,53% até 26,3%. Além disso, para ela, verificou-se que 25% dessas nações apresentam valores de *EDUC* menores que 11,25%. Considerando a porcentagem da população que vive em áreas urbanas, temos que 50% dessas nações apresentam valores inferiores a 60%, com mínimo de 16,10% e máximo de 100%.

Tem-se que cerca de 25% desses países possuem valores de *URB* superiores a 74,82%. O consumo médio de álcool por pessoa em litros apresentou valor mínimo de 0,10 e máximo de 15,40, com média de 7,39. O CV é definido como a razão entre o desvio padrão e a média, sendo classificado como uma medida de dispersão. A partir dele, é possível verificar que a variável *ALC* apresenta a maior variabilidade de dados em relação à média, com CV igual a 0,597. Vale ressaltar que um CV igual a zero nos diz que os dados de uma determinada variável são homogêneos. Ou seja, todas as observações equivalem à média.

O país Colômbia, localizado no continente da América do Sul, apresentou a maior proporção de pessoas que praticam atividade física insuficiente. Outros países estão bem próximo dessa proporção, como Malásia, África do Sul e Maurítânia, sendo o primeiro localizado na Ásia, e os dois últimos, na África. Os maiores valores para a expectativa de vida foram observados na Espanha e Itália, localizados na Europa, seguidos por Cingapura, na Ásia.

O continente europeu se destacou por apresentar o maior consumo de álcool por pessoa. Em ordem decrescente de seus valores, temos Lituânia, Romênia e Hungria. Os países Cingapura, Catar e Bélgica apresentaram as maiores porcentagens de pessoas vivendo em áreas urbanas. Vale ressaltar que os dois primeiros estão localizados na Ásia, e o último, na Europa. O continente da África se destacou por apresentar os maiores gastos com a educação como porcentagem de despesa total pelo governo, a saber, os países Etiópia, Namíbia e Benin. Por fim, a maior proporção de adultos obesos foi observada em Catar, localizado na Ásia, seguido por Estados Unidos, pertencente à América, enquanto que os menores valores foram observados em Camboja e Nepal, localizados no continente asiático.

**Tabela 1**

Dados descritivos das variáveis em estudo.

Variáveis	Mínimo	$Q_{1/4}$	Mediana	Média	$Q_{3/4}$	Máximo	CV
<i>OB2014</i>	0,03	0,07	0,20	0,17	0,26	0,14	0,568
<i>INAT</i>	4,10	18,40	23,80	24,68	30,65	63,60	0,431
<i>VIDA</i>	48,93	65,06	74,41	71,73	79,94	83,08	0,128
<i>EDUC</i>	5,53	11,25	14,36	14,66	17,50	26,30	0,316
<i>URB</i>	16,10	39,22	60,00	57,35	74,82	100,00	0,401
<i>ALC</i>	0,10	3,92	7,15	7,39	11,25	15,40	0,597

CV: coeficiente de variação;  $Q_{1/4}$ : primeiro quartil;  $Q_{3/4}$ : terceiro quartil.

Fonte: dados da pesquisa, 2016.

De acordo com a Tabela 2, pode-se observar que *OB2014* se correlaciona positivamente com a maioria das covariáveis, exceto *EDUC*. Além disso, as maiores correlações lineares com a variável resposta são verificadas para *URB* e *VIDA*. Apesar de haver uma correlação de 0,70, entre elas, não ocorreram problemas relacionados à multicolinearidade na análise de regressão mais adiante.

A Figura 1 apresenta o histograma de frequências e o *boxplot* da variável proporção de adultos obesos em 2014. Nessa figura, é possível observar que a distribuição da variável resposta é assimétrica, facilmente observada no *boxplot*, já a mediana está mais próxima do terceiro quartil. Além disso, é verificado a ausência de *outliers*, ou seja, observações discrepantes que excedem os limites do *boxplot*. Tais limites são definidos a partir das quantidades  $Q_{1/4} - 1,5 \times (Q_{3/4} - Q_{1/4})$  e  $Q_{3/4} + 1,5 \times (Q_{3/4} - Q_{1/4})$ , referindo-se, respectivamente, ao limite inferior e superior.

**Tabela 2**

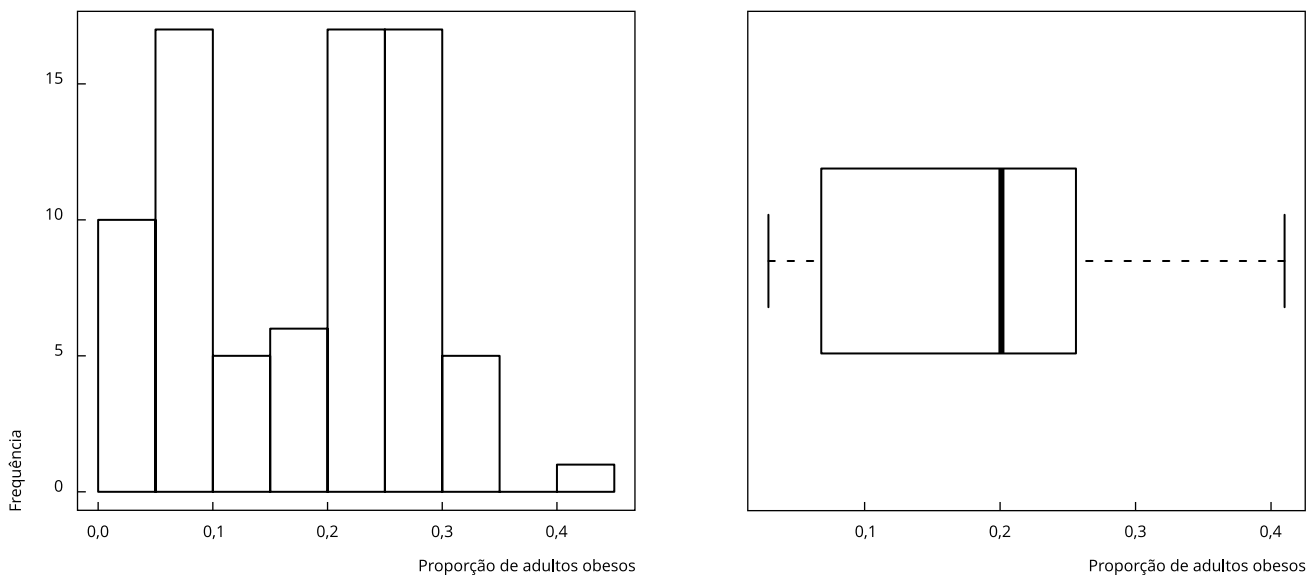
Correlação linear entre as variáveis *OB2014*, *INAT*, *VIDA*, *ALC*, *URB* e *EDUC*.

	<i>OB2014</i>	<i>INAT</i>	<i>VIDA</i>	<i>ALC</i>	<i>URB</i>	<i>EDUC</i>
<i>OB2014</i>	1,00	0,42	0,68	0,57	0,69	-0,29
<i>INAT</i>	-	1,00	0,23	0,05	0,38	-0,03
<i>VIDA</i>	-	-	1,00	0,47	0,70	-0,29
<i>ALC</i>	-	-	-	1,00	0,45	-0,36
<i>URB</i>	-	-	-	-	1,00	-0,22
<i>EDUC</i>	-	-	-	-	-	1,00

Fonte: dados da pesquisa, 2016.

**Figura 1**

Histograma e *boxplot* da variável proporção de adultos obesos nas nações em 2014, respectivamente.



Fonte: dados da pesquisa, 2016.

A Figura 2 apresenta o *boxplot* da variável *OB2014* segundo os continentes África, América, Ásia, Europa e Oceania. Como resultado, pode-se observar que a maior concentração de nações com menores valores de *OB2014* está nos continentes africano e asiático. Por outro lado, os continentes da América, Europa e Oceania apresentam os maiores valores. Vale ressaltar que não existe interseção entre os *boxplots* da Europa e Oceania com os da África e Ásia, significando uma possível diferença existente entre as proporções de adultos obesos nesses continentes.

Para o modelo de regressão beta, foi considerado o conjunto de dados referente à obesidade adulta nas nações que totalizam 78 observações. Inicialmente, ao se ajustar o modelo de regressão beta, é fundamental se questionar a respeito da dispersão dos dados. Modelos de regressão com dispersão variável necessitam de uma estrutura para modelar a precisão dos parâmetros de modo a melhorar os resultados inferenciais <sup>27</sup>.

Para tanto, foi utilizado o teste da razão de verossimilhanças com objetivo de verificar a hipótese nula de precisão fixa, isto é,  $H_0 : \varnothing_1 = \varnothing = \varnothing_n = \varnothing$  <sup>21,25,28</sup>. Como resultado, obteve-se um valor de  $p < 0.0001$  (valor obtido a partir dos dados amostrais e que reflete a probabilidade de rejeitar a hipótese nula dado que ela é verdadeira). Ou seja, considerando o nível de significância de 5%, rejeitamos a hipótese nula de precisão fixa. Portanto, é necessária uma estrutura de regressão para modelar a precisão dos dados.

O modelo de regressão beta com dispersão variável encontra-se apresentado a seguir:

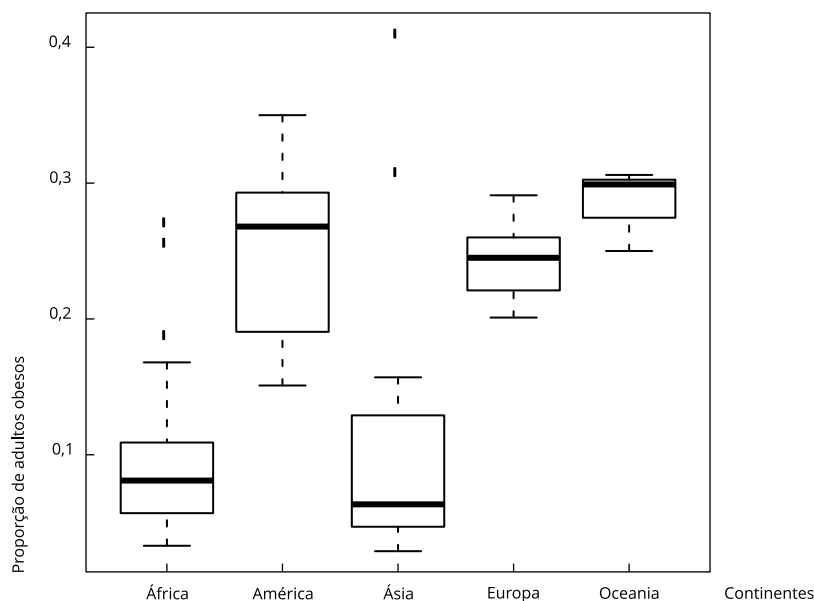
$$\log \log (\mu_t) = \beta_0 + \beta_1 INAT_t + \beta_2 URB_t + \beta_3 ALC_t + \beta_4 VIDA_t$$

$$\log(\varnothing_t) = \gamma_0 + \gamma_1 VIDA_t + \gamma_2 EDUC_2 + \gamma_3 ALC_3$$

com  $t = 1, \dots, 78$ . Nesse modelo, temos que o parâmetro de precisão varia com as observações, havendo assim uma estrutura heteroscedástica. Entretanto, mesmo que a dispersão dos dados seja fixa, a variância da variável resposta não é constante, uma vez que seu valor depende das médias desconhecidas que variam com a estrutura de regressão.

**Figura 2**

*Boxplot* da variável *OB2014* segundo os continentes da África, América, Ásia, Europa e Oceania.



Fonte: dados da pesquisa, 2016.

A Tabela 3 apresenta as estimativas, os erros padrões e os valores de p utilizados para determinar a significância das estimativas do modelo proposto. Aqui, o modelo de regressão beta com dispersão variável utiliza as funções de ligação *loglog* e *log* para relacionar o preditor linear, respectivamente, à resposta média e à precisão. É possível utilizar o teste de Wald <sup>29</sup> para verificar a hipótese nula de que  $\beta_j = 0$  com  $j = 1, \dots, p$ , ou seja, a variável associada ao parâmetro  $\beta_j$  não apresenta efeito significativo sobre a resposta média <sup>30</sup>. Dessa forma, considerando o nível nominal de 5%, temos que as variáveis atividade física insuficiente (*INAT*), pessoas vivendo em áreas urbanas (*URB*), consumo de álcool (*ALC*) e expectativa de vida (*VIDA*) são relevantes para explicar a proporção de adultos obesos nas nações, uma vez que apresentaram valor de  $p < 0,05$ .

Além disso, pode-se destacar que tais covariáveis apresentam efeito positivo no sentido de aumentar a proporção de adultos obesos nas nações. Ou seja, tal resultado se mostra coerente com os obtidos na análise descritiva por meio das correlações lineares com a variável resposta apresentadas na Tabela 2. O efeito positivo da variável *INAT* pode ser justificado pela diminuição da perda de calorías ao longo do dia proporcionada pelas práticas de atividade física insuficiente. Por outro lado, o efeito positivo da variável *URB* pode estar ligado à dificuldade de se realizar refeições em casa devido ao crescente problema na rede de transporte urbano provocado pelo crescimento da urbanização. Dessa forma, a correria da vida moderna incentiva o consumo de refeições fora do domicílio, com destaque para o *fast food* com suas ofertas de alimentos altamente energéticos <sup>31</sup>. Além disso, a modernização e as mudanças no estilo de vida, devido ao avanço tecnológico, tornam as pessoas mais sedentárias e atribuem a elas maiores chances de se tornarem obesas. O efeito positivo da variável *ALC* pode ser entendido a partir da enorme quantidade de calorías ingeridas por meio do consumo do álcool, podendo contribuir para o aumento da obesidade nos países. O processo de envelhecimento das pessoas traz diversas mudanças no corpo como a diminuição do metabolismo e o ganho de peso <sup>33</sup>.

Assim sendo, o efeito positivo da variável *VIDA* pode estar relacionada ao processo de envelhecimento, uma vez que quanto maior a expectativa de vida nas nações maior será a proporção de pessoas em idade mais avançada.

Por exemplo, para as nações com as covariáveis *INAT*, *URB* e *ALC* fixadas na mediana e apresentando uma expectativa de vida de 74 anos, de acordo com o modelo ajustado, estima-se a proporção média de adultos obesos como:

$$\text{loglog}(\mu_t) = -2,009 + 0,009 \times 23,80 + 0,005 \times 60 + 0,027 \times 7,15 + 0,01 \times 74,41$$

Contudo, como a função de ligação utilizada foi a *loglog*, a função inversa aplicada ao preditor linear a fim de obter o valor esperado da variável resposta é

$$\mu_t = \exp \left\{ -\exp (2,009 - 0,009 \times 23,80 - 0,005 \times 60 - 0,27 \times 7,15 - 0,01 \times 74,41) \right\}$$

$$\mu_t = 0.17$$

**Tabela 3**

Estimativas dos coeficientes, erro padrão e valor de p do modelo de regressão beta com dispersão variável, considerando as funções de ligação *loglog* e *log* para modelar a média e a dispersão, respectivamente.

Função de ligação	Variáveis	Parâmetros	Estimativas	Erro padrão	Valor de p
<i>loglog</i> ( $\mu$ )	<i>INT</i>	$\beta_0$	-2,009	0,124	< 0,001
	<i>INAT</i>	$\beta_1$	0,009	0,002	< 0,001
	<i>URB</i>	$\beta_2$	0,005	0,001	< 0,001
	<i>ALC</i>	$\beta_3$	0,027	0,005	< 0,001
	<i>VIDA</i>	$\beta_4$	0,010	0,002	< 0,001
<i>log</i> ( $\phi$ )	<i>INT</i>	$\gamma_0$	9,458	1,546	< 0,001
	<i>VIDA</i>	$\gamma_1$	-0,059	0,020	< 0,001
	<i>EDUC</i>	$\gamma_2$	-0,133	0,036	< 0,001
	<i>ALC</i>	$\gamma_3$	0,099	0,044	0,023

*INT*: refere-se ao intercepto do modelo.

Fonte: dados da pesquisa, 2016.



Ou seja, para as nações com 23,8% de atividade física insuficiente, 60% da população vivendo em áreas urbanas, consumo médio de álcool de 7,15 litros por pessoa e expectativa de vida de 74 anos, é esperada uma proporção de adultos obesos em torno de 0,17 ou 17%.

Em relação à modelagem da precisão, Tabela 3, temos que as covariáveis expectativa de vida (*VIDA*), gastos com a educação pelo governo (*EDUC*) e consumo de álcool (*ALC*) foram estatisticamente relevantes ao nível de 5% de significância. Vale ressaltar que quanto maior forem os valores de *VIDA* e *EDUC* nas nações menor será a precisão dos dados, conseqüentemente, a dispersão aumenta. Por outro lado, quanto maior forem os valores de *ALC* maior será a precisão, ou seja, o aumento da precisão significará uma menor dispersão dos dados, tornando a resposta média mais precisa. Em resumo, modelar a variabilidade dos dados é um artifício que permite melhorar os resultados inferenciais.

Para verificar a qualidade do ajuste do modelo, utilizou-se o coeficiente de determinação ajustado (pseudo- $R^2$ ) e o teste *RESET*<sup>33,34</sup>. O pseudo- $R^2$  é uma medida global da variação explicada e análogo ao coeficiente de determinação utilizado nos modelos de regressão linear. Essa medida é definida como o quadrado do coeficiente de correlação amostral entre  $\eta$  e  $g(y)$ <sup>19</sup>. Dessa forma, com um pseudo- $R^2 = 0,69$ , é dito que as covariáveis são capazes de explicar cerca de 70% da variabilidade total da proporção de adultos obesos nas nações. Além disso, temos que esta medida apresenta valores restritos ao intervalo (0,1), ou seja, quanto mais próximo de um melhor a qualidade do ajuste ou o poder explicativo do modelo.

Para testar a correta especificação do modelo, utilizou-se o teste *RESET* para modelos de regressão beta<sup>21,25,33</sup>. O mecanismo do teste consiste em adicionar como covariável o preditor linear estimado elevado a segunda potência,  $\eta^2$ , ao submodelo da média. A ideia por trás do teste é que se essa covariável tem algum poder em explicar a variável resposta, então, rejeitamos a hipótese nula de ausência de erros de especificação. Ou seja, o modelo proposto apresenta forma funcional correta, e não ocorrem omissões de variáveis<sup>34</sup>. Portanto, com um valor de  $p = 0,0075$ , não temos evidências suficientes para rejeitar a hipótese nula de que o modelo está bem especificado ao nível de 5% de significância.

O gráfico de probabilidade normal com envelope simulado é uma técnica que permite ao investigador identificar desvios na suposição do modelo e possíveis observações discrepantes. Na Figura 3, verifica-se que as observações encontram-se distribuídas de forma aleatória dentro dos limites do envelope e próximo à linha central, apresentando uma quantidade reduzida de observações que excedem levemente esses limites. Portanto, não temos evidências suficientes para discordar da adequação do modelo.

É possível ainda estimar o impacto de uma determinada covariável, a exemplo da porcentagem de atividade física insuficiente sobre a proporção de adultos obesos nas nações da seguinte maneira<sup>22</sup>:

$$\frac{\partial E(y_t)}{\partial INAT_T} = \frac{\partial E(\mu_t)}{\partial INAT_T}$$

em que  $E(.)$  é o valor esperado ou esperança. Ou seja, deriva-se o preditor linear em relação à covariável de interesse a qual se quer estimar o efeito individual.

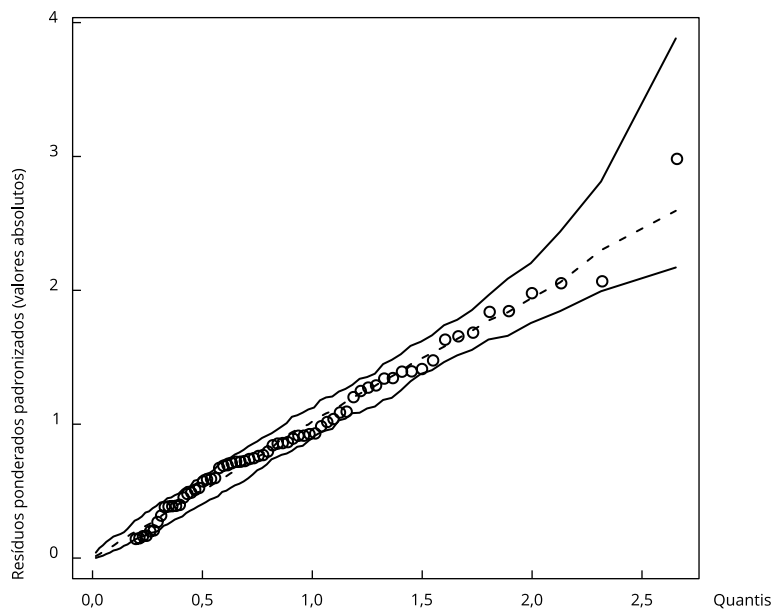
Assim, com o objetivo de estimar as curvas de impacto para descrever o efeito da atividade física insuficiente sobre a proporção de adultos obesos nas nações, foram consideradas três situações, como apresentado na Figura 4. Ou seja, em que as covariáveis *URB*, *ALC* e *VIDA* estão fixadas no primeiro, segundo e terceiro quartis. Dessa forma, é possível variar os valores de *INAT* para determinar o aumento provocado na resposta média. Como resultado, observa-se que o impacto é positivo e cresce lentamente quando se aumentam os valores da atividade física insuficiente. Além disso, não existem grandes diferenças entre as curvas nos quantis 0,50 e 0,75, e elas diminuem à medida que se aumentam os valores de *INAT*. Ou seja, a partir de um determinado valor de *INAT*, próximo à quantidade 0,50, não ocorrem grandes aumentos na resposta média.

## Considerações finais

Diante do exposto, conclui-se que 50% das nações apresentam valores de obesidade maiores que 0,20. Além disso, a expectativa de vida média delas oscila em torno de 72 anos. Vale ressaltar que os valores de atividade física insuficiente são maiores que 23,8% em 50% dos países. A partir da análise do *box-*

**Figura 3**

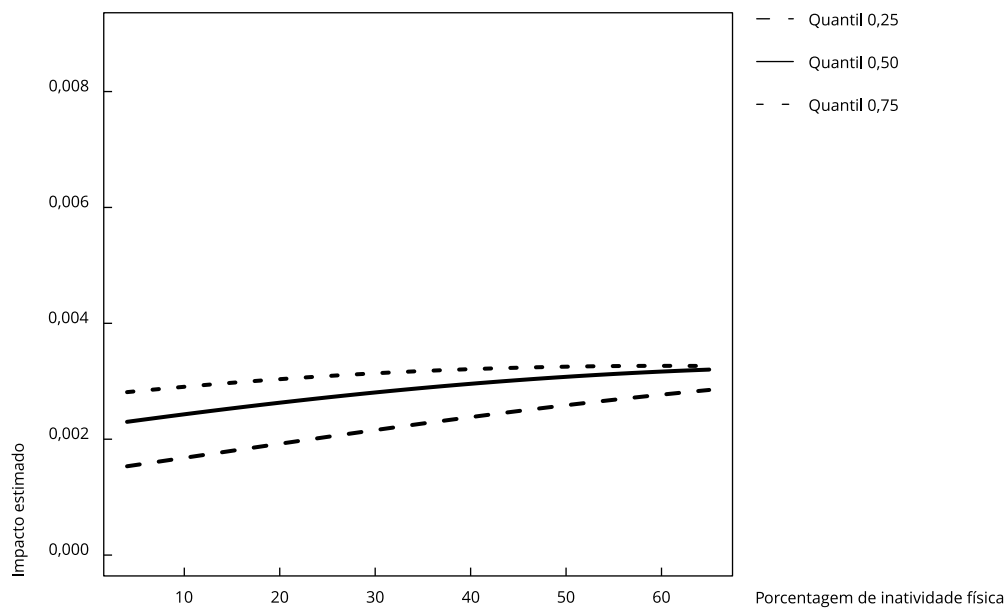
Gráfico da probabilidade normal com envelope simulado.



Fonte: dados da pesquisa, 2016.

**Figura 4**

Impacto da atividade física insuficiente sobre a proporção de obesos nas nações em 2014.



Fonte: dados da pesquisa, 2016.

*plot*, foi observada uma possível diferença nas proporções de adultos obesos entre os continentes da América e Europa com os da África e Ásia.

O modelo de regressão beta utilizado definiu que as covariáveis porcentagem de atividade física insuficiente, porcentagem da população que vive em áreas urbanas, expectativa de vida em anos e o consumo médio de álcool por pessoa em um ano produzem um efeito significativo e positivo sobre a obesidade. Ou seja, elas tendem a aumentar os valores da proporção de adultos obesos quando aumentamos cada uma individualmente enquanto que as demais permanecem constantes.

## Colaboradores

S. A. Souza contribuiu com a concepção do artigo, preparação dos dados, análise estatística e redação do manuscrito. A. B. Silva contribuiu com a redação de trechos do manuscrito e revisão crítica dele. U. M. B. Cavalcante contribuiu com a redação de trechos do manuscrito e revisão crítica dele. C. M. B. L. Lima e T. C. Souza contribuíram com a concepção do artigo, análise dos dados, redação de trechos do manuscrito e revisão crítica dele.

## Agradecimentos

Agradecemos ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pelo apoio financeiro.

## Referências

1. Gigante DP, Dias-da-Costa JS, Olinto MTA, Menezes AMB, Silvia M. Obesidade da população adulta de Pelotas, Rio Grande do Sul, Brasil e associação com nível sócio-econômico. *Cad Saúde Pública* 2006; 22:1873-79.
2. Mariath AB, Grillo LP, Silva RO, Schmitz P, Campos IC, Medina JRP, et al. Obesidade e fatores de risco para o desenvolvimento de doenças crônicas não transmissíveis entre usuários de unidade de alimentação e nutrição. *Cad Saúde Pública* 2007; 23:897-905.
3. Puglia CR. Indicações para o tratamento operatório da obesidade mórbida. *Rev Assoc Méd Bras* 2004; 50:118.
4. Sichieri R, Moura EC. Análise multinível das variações no índice de massa corporal entre adultos, Brasil, 2006. *Rev Saúde Pública* 2009; 43 Suppl 2:90-7.
5. Linhares RS, Horta BL, Gigante DP, Dias-da-Costa JS, Olinto MTA. Distribuição de obesidade geral e abdominal em adultos de uma cidade no Sul do Brasil. *Cad Saúde Pública* 2012; 28:438-47.
6. Carvalho ARM, Belém MO, Oda JY. Sobrepeso e obesidade em alunos de 6-10 anos de escola Estadual de Umuarama/PR. *Arq Ciências Saúde UNIPAR* 2017; 21:3-12.
7. Duncan BB, Chor D, Aquino EML, Bensenor IM, Mill JG, Schmidt MI, et al. Doenças crônicas não transmissíveis no Brasil: prioridade para enfrentamento e investigação. *Rev Saúde Pública* 2012; 46 Suppl 1:126-34.
8. Pinheiro ARO, Freitas SFT, Corso ACT. Uma abordagem epidemiológica da obesidade. *Rev Nutr PUCCAMP* 2004; 17:523-33.
9. Malta DC, Bernal RTI, Andrade SSSA, Silva MMA, Velasquez-Melendez G. Prevalência e fatores associados com hipertensão arterial autorreferida em adultos brasileiros. *Rev Saúde Pública* 2017; 51 Suppl 1:11s.
10. Secretaria de Vigilância em Saúde, Ministério da Saúde. Plano de ações estratégicas para o enfrentamento das doenças crônicas não transmissíveis (DCNT) no Brasil 2011-2022. Brasília: Ministério da Saúde; 2011. (Série B. Textos Básicos de Saúde).
11. Bahia L, Coutinho ESF, Barufaldi LA, Abreu GA, Malhão TA, Souza CPR, et al. The costs of overweight and obesity-related diseases in the Brazilian public health system: cross-sectional study. *BMC Public Health* 2012; 12:440-7.

12. Arterburn D, Maciejewski M, Tsevat J. Impact of morbid obesity on medical expenditures in adults. *Int J Obes (Lond)* 2005; 29:334-9.
13. Organisation for Economic Co-operation and Development. Obesity update, 2014. <http://www.oecd.org/health/Obesity-Update-2014.pdf> (acessado em 30/Jun/2017).
14. Antiporta D, Smeeth L, Gilman RH, Miranda J. Length of urban residence and obesity among within-country rural-to-urban Andean migrants. *Public Health Nutr* 2015; 19:1270-8.
15. Shelton N, Knott C. Association between alcohol calorie intake and overweight and obesity in english adults. *Am J Public Health* 2014; 104:629-31.
16. Gurajati DN, Poter DC. *Econometria básica*. 5ª Ed. Porto Alegre: AMGH Editora; 2011.
17. Kieschnick R, McCullough B. Regression analysis of variates observed on (0,1): percentages, proportions and fractions. *Stat Model* 2003; 3:193-213.
18. Papke L, Wooldridge J. Econometric methods for fractional response variables with an application to 401(k) plan participation rates. *J Appl Econom* 1996; 11:619-32.
19. Ferrari S, Cribari-Neto F. Beta regression for modeling rates and proportions. *J Appl Stat* 2004; 31:799-815.
20. Pereira T. *Regressão beta inflacionada: inferência e aplicações* [Tese de Doutorado]. Recife: Universidade Federal de Pernambuco; 2010.
21. Almeida Junior P, Souza T. Estimativas de votos da presidente Dilma Rousseff nas eleições presidenciais de 2010 sob o âmbito do bolsa família. *Ciênc Nat (Impr)* 2015; 37:12-22.
22. Cribari-Neto F, Souza T. Religious belief and intelligence: worldwide evidence. *Intelligence* 2013; 41:482-9.
23. Espinheira P, Ferrari S, Cribari-Neto F. Influence diagnostics in beta regression. *Computational Statistics & Data Analysis* 2008; 52:4417-31.
24. Espinheira P, Ferrari S, Cribari-Neto F. On beta regression residuals. *J Appl Stat* 2008; 35:407-19.
25. Souza S, Oliveira AA, Souza TC, Lima CMBL. Modelagem da proporção de obesos nos Estados Unidos utilizando modelo de regressão beta com dispersão variável. *Ciênc Nat (Impr)* 2016; 38:1146-56.
26. McCullagh P, Nelder J. *Generalized linear models*. London: Chapman and Hall; 1989.
27. Smithson M, Verkuilen J. A better lemon-squeezer? Maximum likelihood regression with beta-distributed dependent variables. *Psychol Methods* 2006; 11:54-71.
28. Neyman J, Pearson E. On the use and interpretation of certain teste criteria for purposes of statistical inference. *Biometrika* 1928; 20:175-240.
29. Wald A. Test of statistical hypotheses concerning several parameters when the number of observations is large. *Trans Amer Math Soc* 1943; 54:426-82.
30. Cribari-Neto F, Zeileis A. Beta regression in R. *J Stat Softw* 2010; 34:1-24.
31. Anjos LA. *Obesidade e saúde pública*. Rio de Janeiro: Editora Fiocruz; 2006.
32. Souza FR, Schroeder PO, Liberali R. Obesidade e envelhecimento. *Revista Brasileira de Obesidade, Nutrição e Emagrecimento* 2007; 1: 24-35.
33. Lima L. *Um teste de especificação correta para modelos de regressão beta* [Dissertação de Mestrado]. Recife: Universidade Federal de Pernambuco; 2007.
34. Ramsey JB. Tests for specification errors in classical linear least squares regression analysis. *J R Stat Soc* 1969; 31:350-71.

## Abstract

Obesity is considered a serious public health problem, as an epidemic disease with major global repercussions that is associated with the development of other chronic conditions such as hypertension, diabetes, and cardiovascular diseases. The current study examines the distribution of adult obesity in different countries using a beta regression model. This is a descriptive ecological study with a quantitative and inferential approach and a focus on beta regression analysis. Application of this method used a set of real data from public sources on adult obesity in 78 countries in 2014. Descriptive data analysis showed that 50% of the countries showed adult obesity prevalence greater than 20%. In addition, analysis of the distribution of prevalence by country showed lower adult obesity levels in countries of Asia and Africa. Meanwhile, higher values were found in countries of the Americas and Europe. Boxplot analysis also evidenced a possible difference in the proportion of obese adults between the Americas and Europe on one side and Africa and Asia on the other. Adjustment of the beta regression model with varying dispersion and 5% significance identified mean annual per capita alcohol intake, percentage of insufficient physical activity, percentage of the population living in urban areas, and life expectancy as variables associated with adult obesity.

Obesity; Chronic Disease; Linear Models

## Resumen

La obesidad está considerada un grave problema de salud pública, al tratarse de una enfermedad epidémica de gran repercusión en el escenario mundial, que está relacionada con el desarrollo de otras enfermedades crónicas, como, por ejemplo, hipertensión, diabetes y enfermedades cardiovasculares. Ante esto, el presente trabajo tiene como objetivo estudiar la distribución de la obesidad en adultos de diferentes naciones, mediante un modelo de regresión beta. Se trata de un estudio ecológico descriptivo con un abordaje cuantitativo e inferencial, centrándose en el análisis de regresión beta. La aplicación de este método consideró un conjunto de datos reales, obtenidos a partir de fuentes de información pública, referente a la obesidad adulta en las naciones durante el año 2014. Tras el análisis descriptivo de los datos, se verificó que el 50% de las naciones presentan una prevalencia de adultos obesos mayor de un 0,20 (20%). Asimismo, al analizar la distribución de su prevalencia por nación, se constató que los menores valores de obesidad adulta están concentrados en los países pertenecientes a los continentes de Asia y África. Por otro lado, los mayores valores se encuentran distribuidos entre los países en los continentes de América y Europa. Sin embargo, a partir del análisis gráfico del diagrama de caja, se observaron evidencias de una posible diferencia en las proporciones de adultos obesos entre los continentes de América y Europa, respecto a los de África y Asia. Tras ajustar el modelo de regresión beta con dispersión variable, fue posible identificar, con un nivel de un 5% de significancia, que variables como: el consumo medio de alcohol en litros por persona, el porcentaje de actividad física insuficiente, el porcentaje de la población que vive en áreas urbanas y su expectativa de vida presentan efectos en este sentido.

Obesidad; Enfermedad Crónica; Modelos Lineales

---

Recebido em 17/Set/2017

Versão final reapresentada em 13/Mar/2018

Aprovado em 23/Mar/2018