

**MODELAGEM DA OBESIDADE ADULTA NAS NAÇÕES  
 VIA MODELO DE REGRESSÃO QUANTÍLICA**

Saul de Azevedo Souza<sup>1</sup>, Allan Batista Silva<sup>2</sup>  
 Ulanna Maria Bastos Cavalcante<sup>2</sup>, Caliandra Maria Bezerra Luna Lima<sup>2</sup>  
 Tatiene Correia de Souza<sup>2</sup>

**RESUMO**

**Introdução e Objetivo:** A obesidade vem apresentando um importante aumento na sua prevalência e é considerada um grave problema de saúde pública que afeta os países desenvolvidos e em desenvolvimento. O presente trabalho tem como objetivo modelar a obesidade adulta de diferentes nações, por meio do modelo de regressão quantílica. **Metodologia:** Trata-se de um estudo ecológico com abordagem quantitativa. Foram utilizados os dados referentes à proporção de adultos obesos, pessoas maiores de 18 anos com IMC acima de 30kg/m<sup>2</sup>, em 78 países no ano de 2014. As bases consultadas para a coleta dos dados foram *Word Bank* e *World Health Organization*. Os dados foram submetidos ao *software R*, onde foram realizados os procedimentos inferenciais por meio do modelo de regressão quantílica. **Resultados:** Observou-se que as variáveis porcentagem de atividade física insuficiente, porcentagem da população que vivem em áreas urbanas e média do consumo em litros de álcool puro por pessoa em um ano são significativas e apresentam efeito positivo sobre a obesidade adulta nas nações. **Discussão e Conclusão:** Ao estimar o impacto das taxas de inatividade física sobre o desfecho foi possível verificar um impacto positivo e crescente à medida que aumentamos os valores da inatividade física nos diferentes cenários. Além disso, verifica-se que com o modelo de regressão quantílica os gestores possuem mais uma alternativa para tomada de decisão na realização de medidas de controle/prevenção da doença.

**Palavras-chave:** Obesidade. Análise de Regressão. Estudos Epidemiológicos.

1-Programa de pós-graduação em Estatística, Universidade Federal de Pernambuco (UFPE), Recife-PE, Brasil

2-Programa de pós-graduação em Modelos de Decisão e Saúde pela Universidade Federal da Paraíba (UFPB), João Pessoa-PB, Brasil.

**ABSTRACT**

Modeling of adult obesity in the nations via a quantile regression model

**Introduction and Objective:** Obesity has shown a significant increase in its prevalence and is considered a serious public health problem that affects both developed and developing countries. The present work aims to model the adult obesity of different nations, using the quantile regression model. **Methodology:** This is an ecological study with a quantitative approach. Data on the proportion of obese adults, individuals over 18 years of age with a BMI above 30 kg / m<sup>2</sup>, were used in 78 countries in 2014. The databases consulted for data collection were *Word Bank* and *World Health Organization*. The data were submitted to *software R*, where the inferential procedures were performed using the quantile regression model. **Results:** It was observed that the percentages of insufficient physical activity, percentage of the population living in urban areas and average consumption of liters of pure alcohol per person in a year are significant and present a positive effect on adult obesity in the nations. **Discussion and Conclusion:** When estimating the impact of physical inactivity rates on the outcome, it was possible to verify a positive and increasing impact as we increase the values of physical inactivity in the different scenarios. In addition, it is verified that with the quantile regression model managers have another alternative for decision-making in the accomplishment of measures of control / prevention of the disease.

**Key words:** Obesity. Regression analysis. Epidemiologic studies.

E-mails:

saul\_asouza@hotmail.com

allandobu@gmail.com

ulannacavalcante@hotmail.com

calilunlima@gmail.com

tatitex@gmail.com

## INTRODUÇÃO

A obesidade se caracteriza pelo acúmulo excessivo de gordura no organismo, comprometendo a saúde dos indivíduos, deixando-os vulneráveis para o surgimento de outras doenças crônicas não transmissíveis (Pereira e colaboradores, 2014).

Anualmente cerca de 3,4 milhões de pessoas no mundo morrem por conta do excesso de peso e obesidade. Além disso, também a nível mundial, essas condições são responsáveis por 44% dos diagnósticos de diabetes, 23% de doenças isquêmicas do coração e entre 7% e 41% de alguns tipos de câncer (Lima e colaboradores, 2015).

A obesidade vem apresentando um importante aumento na sua prevalência e é considerada um grave problema de saúde pública que afeta os países desenvolvidos e em desenvolvimento, independente da renda, sexo e idade (Aranceta-Bartrina e colaboradores, 2016; Cordeiro e colaboradores, 2016; Salas e colaboradores, 2017).

Entre os anos de 1980 e 2003 a prevalência global do excesso de peso e a obesidade aumentaram cerca de 27,5% na população adulta, sendo que nos homens esse aumentou foi de 28,8% em 1980 para 36,9% em 2013, já nas mulheres o incremento observado foi de 29,8% para 38% (NG e colaboradores, 2014).

Alguns autores sugerem que o aumento da obesidade ou sobrepeso no mundo deve-se a transição de uma alimentação saudável para um consumo exagerado de alimentos altamente energéticos, destacando uma dieta mais rica em gorduras animais, açúcares ou lipídeos, favorecendo o aumento da adiposidade. Além disso, a falta de incentivos a práticas de atividade física, muitas vezes ocasionada pelo avanço tecnológico ou mudanças no estilo de vida, atribuem as pessoas maiores chances de se tornarem obesas (Pereira e colaboradores, 2014; Lima e colaboradores, 2015; Garcia e colaboradores, 2016; Carvalho, Belém, Oca, 2017).

Por ser um método simples, o IMC – índice de Massa Corporal é o parâmetro mais utilizado para avaliar e categorizar os indivíduos com base na concentração de gordura. Esse parâmetro é obtido pelo resultado da razão entre o peso do indivíduo dado em quilogramas (kg) e sua altura ao quadrado (m<sup>2</sup>).

Segundo a Organização Mundial da Saúde (OMS), o IMC é classificado nas seguintes categorias: baixo peso (IMC < 18,5 kg/m<sup>2</sup>), eutrófico (IMC entre 18,5 e 24,9 kg/m<sup>2</sup>), sobrepeso (IMC entre 25 e 29,9 kg/m<sup>2</sup>) e obeso (IMC ≥ 30 kg/m<sup>2</sup>) (Linhares e colaboradores, 2012; Garcia e colaboradores 2016).

O IMC elevado é um fator de risco para a morbimortalidade de várias doenças, principalmente as doenças cardiovasculares, cujas são a principal causa de morte em todo o mundo (Zapata, Bibiloni, Tur, 2016; Callo e colaboradores, 2016).

A obesidade é tida como um importante problema mundial de saúde pública, cuja gera impactos negativos na vida dos indivíduos e de seus familiares. Por esse motivo os estudos que buscam compreender a prevalência e os possíveis fatores de risco da obesidade são de grande relevância, pois possibilitam uma melhor adequação e adaptação das medidas preventivas aplicadas. (Amaral, Pereira, 2008).

Dessa forma, o presente trabalho tem como objetivo avaliar a distribuição da obesidade em adultos de diferentes nações, por meio do modelo de regressão quantílica.

## MATERIAIS E MÉTODOS

### Tipo de estudo e Descrição dos dados

O presente trabalho consiste num recorte da dissertação de Souza (2017). Trata-se de um estudo ecológico, do tipo descritivo quanto ao objetivo e com abordagem quantitativa.

Foram utilizados os dados referentes a proporção de obesos adultos, pessoas maiores de 18 anos com IMC acima de 30kg/m<sup>2</sup>, em 78 países no ano de 2014.

A amostra, constituída pelas 78 observações (proporções) referentes às nações no mundo é distribuída da seguinte maneira: 25 países (32%) pertencem a África, 11 (14%) a América, 14 (18%) a Ásia, 25 (32%) a Europa e 3 (4%) a Oceania.

As bases consultadas para a coleta dos dados foram Word Bank (<http://databank.worldbank.org>) e World Health Organization (<http://www.who.int>). Neste estudo a variável dependente observada foi a Proporção de adultos obesos (OB2014), 18 anos ou mais, com IMC ≥ 30 kg/m<sup>2</sup> em 2014.

Por outro lado, as variáveis explicativas foram: Porcentagem de atividade

física insuficiente entre adultos em 2010 (INAT), Gastos com a educação como porcentagem da despesa total do governo em 2010 (EDUC), Expectativa de vida ao nascer em anos no ano de 2014 (VIDA), Média do consumo em litros de álcool puro por pessoa em um ano, considerando a população com 15 anos ou mais em 2008 (ALC) e Porcentagem da população que vivem em áreas urbanas em 2014 (URB).

Vale ressaltar que foram coletados os dados mais atualizados que estavam disponíveis e que compreendia um quantitativo de nações. E por se tratar de dados de domínio público, não se faz necessário o envio da pesquisa para Comitê de Ética em Pesquisa.

### Procedimento de Análise dos Dados

Os dados foram tabulados em planilha eletrônica e submetidos ao *software R* (R Core Team, 2013).

Inicialmente, foi realizada uma análise descritiva das covariáveis abordadas no estudo. Adicionalmente, procedeu-se com a construção de mapas referentes as variáveis em estudo com o objetivo de visualizar os níveis destas covariáveis nos diferentes países.

Em seguida, foram realizados os procedimentos inferenciais por meio do modelo de regressão quantílica.

Vale ressaltar que para tal modelo utilizou-se os pacotes *quantreg* (Koenker, 2005), disponíveis no *software R*.

### Modelo de Regressão Quantílica

O modelo de regressão quantílica proposto por Koenker e Bassett (1978) trata de uma abordagem mais geral que os modelos clássicos de regressão e pode ser caracterizado como um modelo de regressão semiparamétrico, pois não exige nenhuma

distribuição de probabilidade para a variável resposta.

Algumas características úteis da regressão quantílica podem ser sintetizadas a seguir: tal abordagem pode ser utilizada para caracterizar toda a distribuição condicional da variável resposta dado um conjunto de variáveis independentes; essa classe de modelos têm uma representação de programação linear que torna as estimativas dos parâmetros mais simples; o seu método de estimação é responsável por atribuir um caráter mais robusto a regressão quantílica, de modo que a estimativa dos parâmetros não seja sensível a outliers na variável resposta; quando os erros não são normais, seu estimador pode ser mais eficiente do que o estimador de mínimos quadrados ordinários; como as estimativas dos parâmetros em cada quantil são obtidas de forma independente é possível obter soluções potencialmente diferentes modificando as covariáveis nos diferentes pontos da distribuição condicional da variável resposta (Buchinsky, 1998).

A regressão quantílica ainda apresenta uma característica distinta denominada de equivariância a transformações monótonas (Bottai, Cai, Mckeown, 2010; Hao, Naiman, 2007; Mua; Hea, 2007; Powell, 1986; Santos, 2012).

Ou seja, caso torne-se necessário aplicar uma função  $h(\cdot)$ , a exemplo da exponencial ou da logarítmica, não decrescente no conjunto  $R$  na variável resposta  $y$ , temos que os quantis serão obtidos por aplicar a mesma transformação na função quantílica,  $Q_{\tau}(y|x)$ .

Por exemplo, se  $q$  é o  $\tau$ -ésimo quantil de  $y$ , então  $h(q)$  é o  $\tau$ -ésimo quantil de  $h(y)$ .

Portanto, considere  $y$  uma variável resposta restrita ao intervalo  $(0,1]$ , com distribuição de probabilidade não gaussiana, e  $x$  um vetor de covariáveis independentes.

Dessa forma, é possível estabelecer uma relação funcional entre a variável  $y$  e o vetor de covariáveis,  $x$ , da seguinte maneira:

$$h\{Q_{\tau}(y_t|x_t)\} = \beta_{0(\tau)} + \beta_{1(\tau)}x_{t1} + \dots + \beta_{p(\tau)}x_{tp}, \quad (1)$$

em que  $Q_{\tau}(y_t|x_t)$  é o quantil condicional de  $y|x$  e  $\beta_{0(\tau)}, \dots, \beta_{p(\tau)}$  são os parâmetros desconhecidos indexados no quantil de ordem  $\tau$ , com  $\tau \in (0,1)$ . O quantil de ordem  $\tau$  pode ser definido como um ponto em que 100 $\tau$ % dos valores amostrais ou populacionais são menores do que ele. Aqui, temos que  $h(\cdot)$  é

uma função estritamente monótona e não decrescente no conjunto real. Existem algumas escolhas possíveis para as funções de ligação  $h(\cdot)$ , a exemplo das ligações probit, loglog e logit (McCullagh, Nelder, 1989).

Portanto, uma vez estimados os coeficientes  $\beta_{j(\tau)}$  é possível obter os valores de

$Q_{\tau}(y_t|x_t)$  aplicando uma transformação inversa  $h^{-1}(\cdot)$  em (1). Por exemplo, podemos utilizar a função  $h(y) = \text{logit}(y)$  para transformar a

variável resposta  $y$ , restrita ao intervalo  $(0, 1]$ , da seguinte maneira:

$$h(y) = \text{logit}(y) = \log\left(\frac{y + \epsilon}{1 - y + \epsilon}\right),$$

em que  $\epsilon$  é uma pequena quantidade definida para que a transformação seja aplicada em todo o conjunto das observações (Bottai, Cai, Mckeown, 2010).

Dessa forma, a inferência a respeito do  $\tau$ -ésimo quantil da variável resposta original pode ser obtida por meio da seguinte transformação inversa, definida como:

$$Q_{\tau}(y_t|x_t) = \frac{\exp(\hat{\beta}_{0(\tau)} + \hat{\beta}_{1(\tau)}x_{t1} + \dots + \hat{\beta}_{p(\tau)}x_{tp})}{1 + \exp(\hat{\beta}_{0(\tau)} + \hat{\beta}_{1(\tau)}x_{t1} + \dots + \hat{\beta}_{p(\tau)}x_{tp})},$$

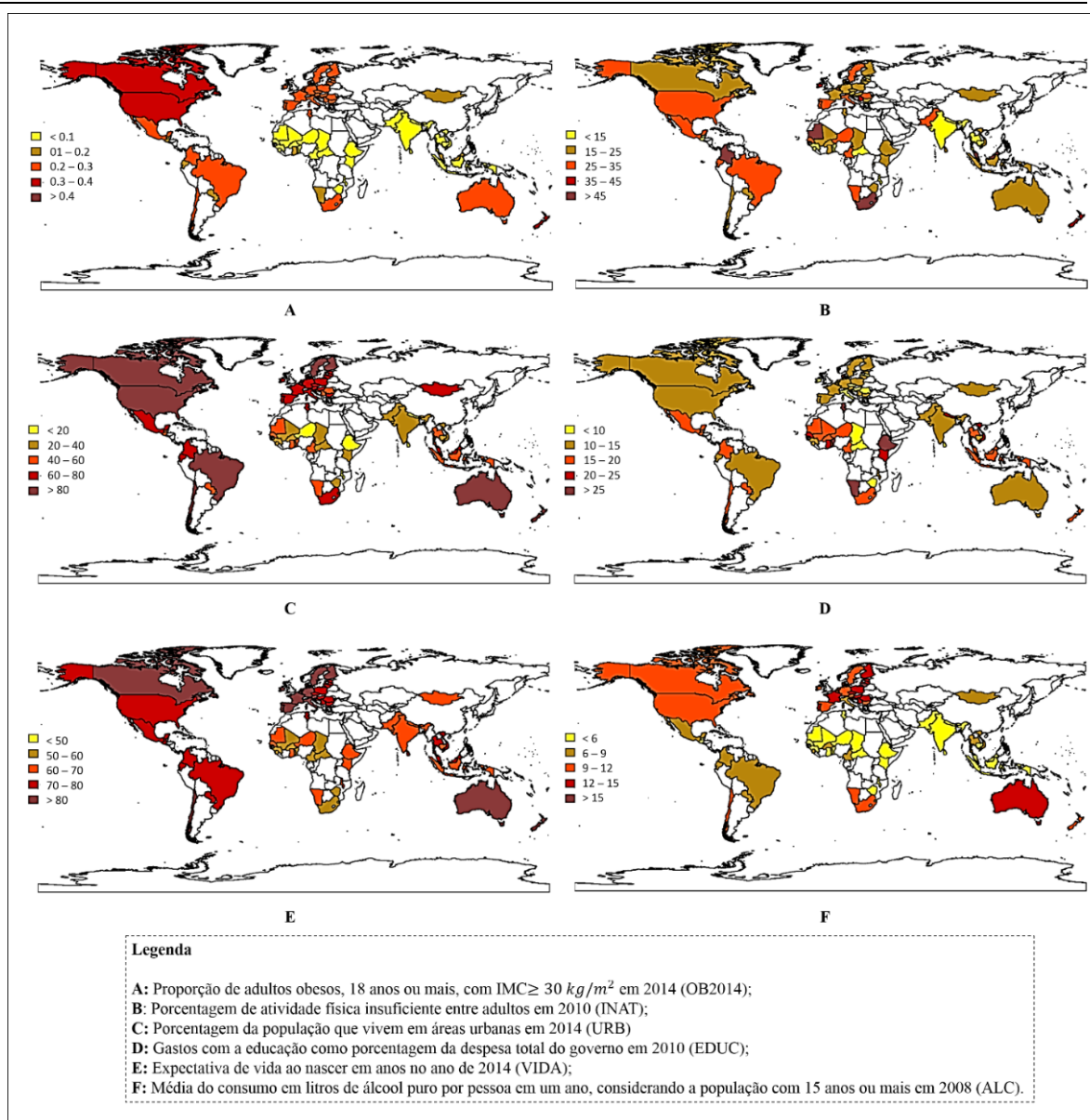
em que os parâmetros desconhecidos  $\hat{\beta}$  são estimados a partir da solução de um problema de minimização como descrito em Neyman, Pearson (1928).

## RESULTADOS E DISCUSSÃO

A Figura 1 apresenta os mapas de OB2014, INAT, URB, EDUC, VIDA E ALC, afim de visualizar de forma mais geral a distribuição dos valores destas covariáveis por nação ou continente. Por exemplo, em relação

ao mapa da obesidade (Figura 1A), as regiões amareladas referem-se as nações com menores proporções de adultos obesos, ou seja,  $OB2014 < 0,1$ .

Por outro lado, as regiões alaranjadas e avermelhadas referem-se as nações com valores mais altos da variável desfecho. Portanto, as nações com menores prevalências de adultos obesos estão distribuídas no continente da África e Ásia, enquanto as nações com maiores valores encontram-se na América, Europa e Oceania.



**Figura 1 -** Mapas das nações segundo as variáveis abordadas no estudo.

O modelo de regressão quantílica foi utilizado para explicar a relação existente entre a proporção de adultos obesos nas nações e as demais covariáveis.

O modelo de regressão quantílica proposto por Koenker e Bassett (1978) é caracterizado como uma abordagem mais geral que os modelos clássicos, pois permite ajustar retas de regressão para diversos quantis de interesse e tratar dos problemas relacionados à assimetria nos dados.

Além disso, este método apresenta uma propriedade básica denominada de equivariância a transformações monótonas, como apresentada em Bottai, Cai e McKeown (2010), Hao e Naiman (2007), Santos (2012).

Na literatura existem várias possibilidades de transformação para variáveis restritas ao intervalo (0,1).

Contudo a transformação logit é a mais utilizada. Assim, a variável transformada pode ser definida da seguinte maneira:

$$h(y) = \text{logit}(OB2014) = \log\left(\frac{OB2014}{1 - OB2014}\right),$$

em que  $h(\cdot)$  é a função logit. Essa transformação permite que a variável assumira valores no conjunto dos números reais, possibilitando assim o uso da regressão quantílica.

O uso dessa abordagem é denominado de regressão quantílica logística como apresentado em Bottai, Cai e McKeown (2010), Luca e Boccuzzo (2014), Feizi, Aliyari e Roohafza (2012).

Esta propriedade da regressão quantílica não é aplicável a regressão das médias. Ou seja, não é possível ajustar um modelo de regressão normal linear com a variável  $h(y)$  e esperar que a mesma

interpretação seja obtida em  $y$ , após aplicar a função inversa  $h^{-1}$ .

Os procedimentos computacionais foram realizados utilizando o pacote *quantreg* (Koenker, 2005; Koenker, 2008) do software estatístico R (R Core Team, 2013), que é uma plataforma livre e apresenta em suas bibliotecas diversos métodos estatísticos já implementados.

A seguir está apresentado o modelo de regressão quantílica utilizado para obter informações a respeito do  $\tau$ -ésimo quantil condicional do logit da proporção de adultos obesos nas nações.

$$Q_{\tau}(\text{logit}(OB2014_t|x_t) = \beta_{0(\tau)} + \beta_{1(\tau)}INAT + \beta_{2(\tau)}URB + \beta_{3(\tau)}ALC,$$

com  $t = 1, \dots, 78$  e parâmetros indexados no quantil  $\tau$ , com  $\tau \in (0,1)$ .

A Tabela 1 apresenta as estimativas dos coeficientes após ajustar o modelo de regressão quantílica para  $\tau = 0,15$ ,  $\tau = 0,25$ ,  $\tau = 0,50$ ,  $\tau = 0,75$ ,  $\tau = 0,85$  e  $\tau = 0,90$ , além da regressão normal linear referenciada por MQO (Mínimos Quadrados Ordinários).

Vale ressaltar que na literatura não existe um valor fixo de  $\tau$  para ser avaliado.

Dessa forma a escolha desses quantis busca analisar o efeito das covariáveis em diferentes pontos na cauda inferior, superior e no quantil mediano da distribuição condicional da variável resposta.

Aqui, o  $p$ -valor é obtido assumindo que os resíduos são não identicamente distribuídos tornando as análises mais robustas.

A partir dos resultados é possível observar que a variável porcentagem de indivíduos que praticam atividade física insuficiente (INAT) apresenta efeito positivo para todos os quantis avaliados.

Ou seja, quanto menor for os valores de INAT menor será a prevalência de adultos obesos nas nações (mantendo as demais covariáveis constantes). Isto ocorre por que as práticas de atividade física se apresentam como um fator de proteção para a obesidade, reduzindo o acúmulo de calorias ao longo do dia (Amaral, Pereira, 2008).

Entretanto em alguns pontos da distribuição condicional da variável resposta está covariável não demonstrou ser relevante dado o  $p$ -valor  $> 0,05$ .

As estimativas dos parâmetros relacionados a variável porcentagem de pessoas vivendo em áreas urbanas (URB) revela um efeito positivo sobre a resposta. Este resultado pode ser devido as mudanças no estilo de vida proporcionada pela crescente urbanização, uma vez que a correria da vida moderna incentiva o consumo de refeições fora do domicílio. Ou seja, as pessoas tornam-se mais propensas a ingerir alimentos mais calóricos, contribuindo para o acúmulo de gordura (Anjos, 2006).

Contudo esta covariável não demonstrou efeito significativo nos quantis 0,85 e 0,90 dado o  $p$ -valor  $> 0,05$ . A variável consumo médio de álcool por pessoa (ALC) também apresentou efeito positivo sobre a prevalência de obesos nas nações. Este resultado pode ser explicado a partir da ingestão de calorias provenientes do álcool que pode contribuir para o aumento da obesidade. Além disso, esta covariável se mostrou relevante para todos os quantis analisados e seu efeito parece diminuir em função dos quantis.

Analisando o ajuste proveniente da regressão normal linear, referenciado por MQO (Mínimos Quadrados Ordinários) Tabela 1, é verificado que o efeito médio destas covariáveis é significativo para explicar a variabilidade da proporção de adultos obesos nas nações.

Além disso, comparando as estimativas de MQO com as estimativas no quantil mediano é possível observar algumas diferenças.

**Tabela 1** - Estimativa dos coeficientes e p-valor referente ao  $\tau$ -ésimo quantil,  $\tau = 0,15; 0,25; 0,50; 0,75; 0,85; 0,90$ ; e MQO. p-valor em parêntese e erro padrão assumido ser não identicamente distribuído (n.i.d).

| Variáveis | Parâmetros | Quantis     |             |             |             |             |             | MQO    |
|-----------|------------|-------------|-------------|-------------|-------------|-------------|-------------|--------|
|           |            | $\tau=0,15$ | $\tau=0,25$ | $\tau=0,50$ | $\tau=0,75$ | $\tau=0,85$ | $\tau=0,90$ |        |
| INT       | $\beta_0$  | -4,375      | -4,222      | -3,680      | -3,774      | -2,727      | -2,617      | -3,816 |
|           |            | <0,001      | <0,001      | <0,001      | <0,001      | <0,001      | <0,001      | <0,001 |
| INAT      | $\beta_1$  | 0,013       | 0,014       | 0,010       | 0,022       | 0,028       | 0,037       | 0,021  |
|           |            | 0,385       | 0,001       | 0,118       | 0,077       | 0,011       | <0,001      | <0,001 |
| URB       | $\beta_2$  | 0,023       | 0,019       | 0,017       | 0,022       | 0,011       | 0,006       | 0,017  |
|           |            | <0,001      | <0,001      | <0,001      | <0,001      | 0,062       | 0,133       | <0,001 |
| ALC       | $\beta_3$  | 0,076       | 0,096       | 0,091       | 0,058       | 0,033       | 0,043       | 0,075  |
|           |            | 0,013       | <0,001      | <0,001      | 0,013       | 0,077       | 0,008       | <0,001 |

**Legenda:** INT: Refere-se ao intercepto do modelo.

Por exemplo, a estimativa de  $\beta_1$  via MQO é quase duas vezes maior que a estimativa no quantil mediano para mesma covariável.

Contudo, a significância destas diferenças será determinada a partir de análise gráfica mais adiante.

Dessa forma, para países com 23,80% de atividade física insuficiente, 60% da população vivendo em áreas urbanas e consumo médio de álcool equivalente a 7,15 litros por pessoa, de acordo com o modelo ajustado, estima-se que o quantil de ordem 0.50 do logit da proporção de adultos obesos é

$$Q_{0,50}(\text{logit}(\text{OB2014})|X) = -3,680 + 0,010 \times 23,80 + 0,017 \times 60 + 0,091 \times 7,15$$

Contudo, pela propriedade de equivariância a transformações monótonas temos que

$$Q_{0,50}(\text{logit}(\text{OB2014})|X) = \frac{\exp(-3,680 + 0,010 \times 23,80 + 0,017 \times 60 + 0,091 \times 7,15)}{1 + \exp(-3,680 + 0,010 \times 23,80 + 0,017 \times 60 + 0,091 \times 7,15)} \approx 0,15$$

Ou seja, para as nações com 23,80% de atividade física insuficiente, 60% da população vivendo em áreas urbanas e consumo médio de álcool equivalente a 7,15 litros por pessoa é esperado uma proporção de adultos obesos em torno de 0,15 ou 15%.

A Figura 2 permite visualizar as estimativas dos parâmetros da regressão quantílica e compará-las com as que seriam obtidas caso se ajustasse um modelo de regressão normal linear em função dos quantis. Ou seja, é possível visualizar de forma mais geral o comportamento das estimativas de cada parâmetro em diferentes quantis da distribuição condicional da variável resposta. Para comparação, as linhas pontilhadas são as estimativas e o intervalo com 95% de confiança do estimador de mínimos quadrados ordinários (MQO).

Por outro lado, a área hachurada refere-se ao intervalo com 95% de confiança das estimativas da regressão quantílica.

Dessa forma, para obter um gráfico mais detalhado definiu-se um conjunto de quantis mais amplo. Como resultado é verificado que as estimativas do intercepto,

denotado por INT, crescem gradativamente em função dos quantis.

Além disso,  $\hat{\beta}_{0(\tau)}$  demonstra ser significativo para todos os quantis avaliados, uma vez que o valor zero não está contido no intervalo com 95% de confiança da regressão quantílica.

As estimativas dos coeficientes relacionado a INAT oscilam dentro do intervalo de confiança via MQO em boa parte dos quantis, especificamente nos quantis inferiores e mediano. É observado também que suas estimativas crescem lentamente para  $\tau > 0,50$ .

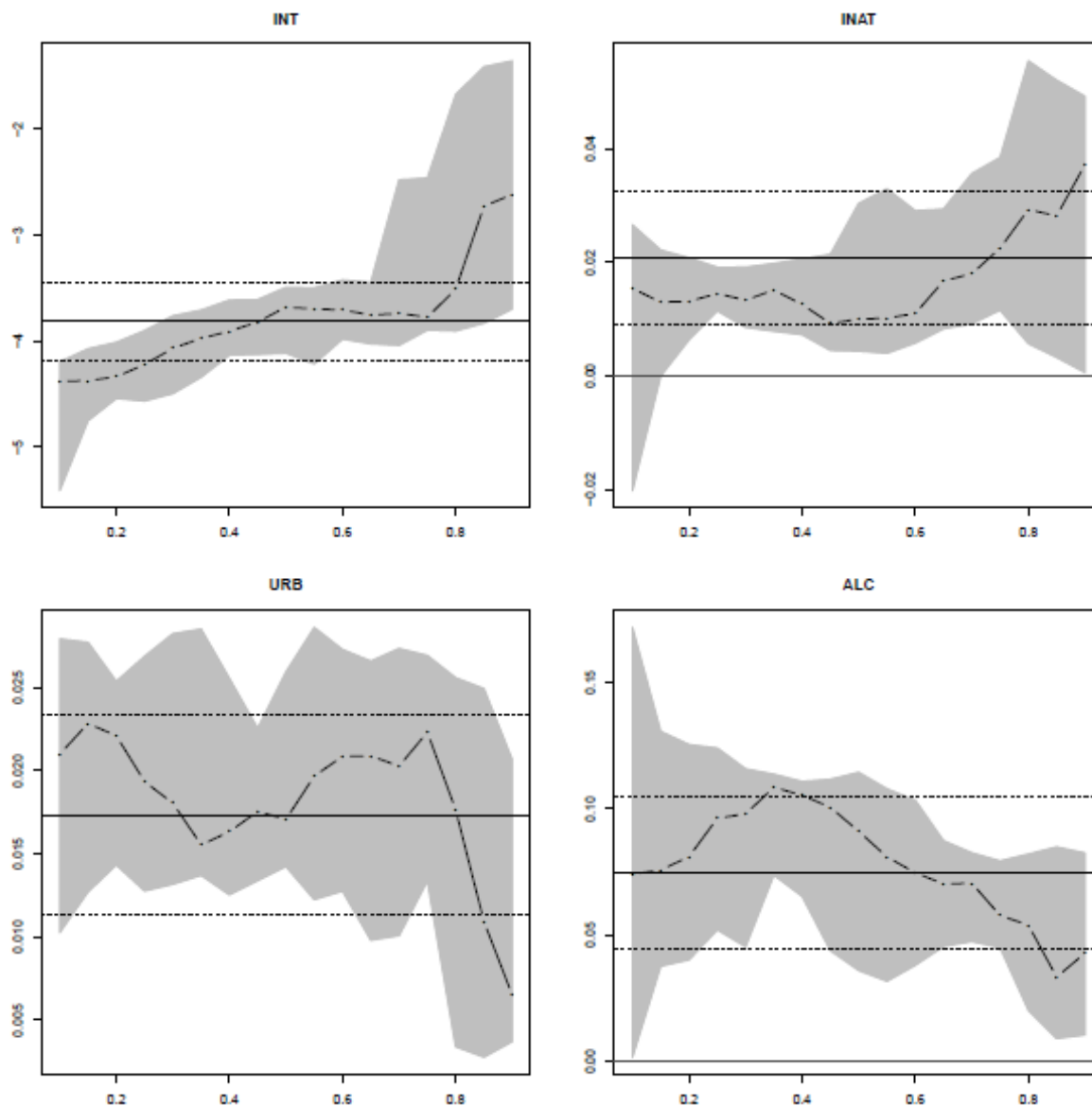
As estimativas relacionadas a variável URB apresentam um comportamento oscilatório com um decréscimo acentuado a partir do quantil 0,75.

Além disso, a maior parte de suas estimativas, especificamente para  $\tau < 0,75$ , está contida no intervalo via MQO.

As estimativas relacionadas a variável ALC assumem um padrão decrescente para o quantis maiores que 0,40. É visto também que em diversos pontos as estimativas da regressão quantílica se mantêm contidas dentro do intervalo de confiança de mínimos

quadrados ordinários, não demonstrado haver diferença estatística entre as estimativas

obtidas por meio das duas abordagens.



**Figura 2** - Estimativas (linhas contínuas) e intervalo de confiança de 95% (área hachurada) para os coeficientes de regressão considerando um conjunto denso de quantis,  $\tau = 0,05; 0,10; \dots; 0,90$ . A linha horizontal em zero é marcada como referência. As linhas pontilhadas referem-se ao intervalo de confiança de mínimos quadrados ordinários.

Após estimar os parâmetros dos modelos um ponto importante é verificar se as diferenças observadas nos diferentes quantis é realmente significativa. Para isso, pode-se utilizar o teste de Wald (Buchinsky, 1998; Koenker, Bassett, 1982; Neyman, Pearson, 1928) cuja hipótese nula é  $H_0: \beta_{2(\tau)} = \beta_{2(\theta)}$  e  $\beta_{3(\tau)} = \beta_{3(\theta)}$  e ... e  $\beta_{j(\tau)} = \beta_{j(\theta)}$  versus a alternativa  $H_1: \beta_{2(\tau)} \neq \beta_{2(\theta)}$  ou  $\beta_{3(\tau)} \neq \beta_{3(\theta)}$

$\beta_{3(\theta)}$  ou ... ou  $\beta_{j(\tau)} \neq \beta_{j(\theta)}$ , com  $\theta \neq \tau$  e  $\theta, \tau \in (0,1)$ . Ou seja, é possível testar de forma simultânea múltiplos coeficientes e verificar se a função do  $\tau$ -ésimo e  $\theta$ -ésimo quantil condicional são diferentes uma da outra. Além disso, este teste permite verificar a real necessidade de se utilizar a regressão quantílica visto que não é preciso ajustar diferentes modelos se o efeito das covariáveis



é uniforme ao longo dos quantis (Santos, 2012).

Como resultado obteve-se um p-valor < 0,01 ao testar todos os parâmetros de regressão nos quantis  $\tau = (0,15; 0,25; 0,50; 0,75; 0,90)$  de forma simultânea.

Portanto, a rejeição da hipótese nula sugere que pelo menos um dos coeficientes é estatisticamente diferente dos demais, justificando o uso da regressão quantílica.

É possível ainda aplicar o teste de Wald individualmente para cada parâmetro e suas hipóteses são definidas da seguinte maneira  $H_0: \beta_{j(\tau)} = \beta_{j(\theta)}$  versus a alternativa  $H_1: \beta_{j(\tau)} \neq \beta_{j(\theta)}$ .

Ou seja, a hipótese nula sugere que a variável associada ao parâmetro  $\beta_j$  apresenta efeito uniforme nos diferentes quantis especificados.

A Tabela 8 contém os p-valores obtidos após comparar a estimativa atual (por exemplo, no quantil 0,15) contra as demais estimativas referentes ao primeiro quartil, ao mediano, ao terceiro quartil e a posição no quantil 0,90.

Como resultado, é observado que o efeito da variável URB no quantil 0,90 é significativamente diferente daqueles nos quantis 0,15, 0,25, 0,50 e 0,75, considerando o nível nominal de 5%. Por outro lado, O efeito das variáveis INAT e ALC nos quantis 0,25 e 0,50 mostraram diferenças significativas quando comparadas com as estimativas no quantil 0,90.

Portanto, o efeito destas covariáveis sobre a proporção de adultos obesos é positivo e pode variar dependendo da posição observada na distribuição condicional da variável resposta. Especificamente, quando comparado o efeito exercido nos quantis inferiores e mediano com o efeito no quantil superior, ou seja, na posição 0,90.

Com o objetivo de avaliar a qualidade do ajuste dos modelos utilizou-se o teste da falta de ajuste para verificar a hipótese nula de linearidade do modelo contra a hipótese alternativa de não linearidade (He, Zhu, 2003).

Outra maneira de verificar a qualidade do ajuste é a partir de uma medida similar ao pseudo- $R^2$ , introduzida por Koenker e Machado (1999), denotada por  $R^1(\tau)$ .

Na Tabela 2 pode-se verificar que apenas o modelo ajustado no quantil mediano rejeitou a hipótese de linearidade, uma vez que seu p-valor é menor do que o nível de significância de 5%.

Em relação a medida da bondade de ajuste,  $R^1(\tau)$ , é possível notar um decréscimo dos seus valores ao longo dos quantis, nos levando a entender que a contribuição das variáveis independentes na distribuição condicional da variável resposta não é a mesma.

Ou seja, suas contribuições são maiores na cauda inferior da distribuição. Vale ressaltar que a maior contribuição é avaliada no quantil 0,25 apresentando um valor de 0,55.

**Tabela 2 - Medida da bondade de ajuste e p-valor referente ao teste da falta de ajuste.**

| $\tau$ - étimo quantil | Medida da bondade de ajuste | p-valor |
|------------------------|-----------------------------|---------|
| 0,15                   | 0,52                        | 0,075   |
| 0,25                   | 0,55                        | 0,143   |
| 0,50                   | 0,52                        | 0,004   |
| 0,75                   | 0,37                        | 0,088   |
| 0,85                   | 0,20                        | 0,045   |
| 0,90                   | 0,23                        | 0,354   |

Por fim, é possível computar o efeito marginal de uma determinada covariável por meio de suas respectivas derivadas parciais (Koenker, 2005; Geraci, 2016). Ou seja,

$$Q_{\tau}(h(Y)|X) = X\beta_j$$

em que  $Y$  é a variável resposta,  $X$  é a matriz das covariáveis independentes e  $h(\cdot)$  é uma

considere a função do  $\tau$ -ésimo quantil condicional dada por

transformação não decrescente no conjunto dos reais com inversa  $h^{-1}$ .

Dessa forma, o efeito marginal pode

ser obtido da seguinte maneira

$$\frac{\partial Q_{\tau}(Y|X)}{\partial x_j} = \frac{\partial h^{-1}\{Q_{\tau}(h(Y)|X)\}}{\partial x_j},$$

em que  $x_j$  é a  $j$ -ésima covariável com relação ao qual o efeito marginal deve ser calculado.

Portanto, como a transformação logit foi utilizada para modelar a proporção de adultos obesos a inversa,  $h^{-1}$ , é definida como

$$h^{-1}\{Q_{\tau}(\text{logit}(y)|X)\} = \frac{\exp(\hat{\beta}_{0(\tau)} + \hat{\beta}_{1(\tau)}\text{INAT} + \hat{\beta}_{2(\tau)}\text{URB} + \hat{\beta}_{3(\tau)}\text{ALC})}{1 + \exp(\hat{\beta}_{0(\tau)} + \hat{\beta}_{1(\tau)}\text{INAT} + \hat{\beta}_{2(\tau)}\text{URB} + \hat{\beta}_{3(\tau)}\text{ALC})}$$

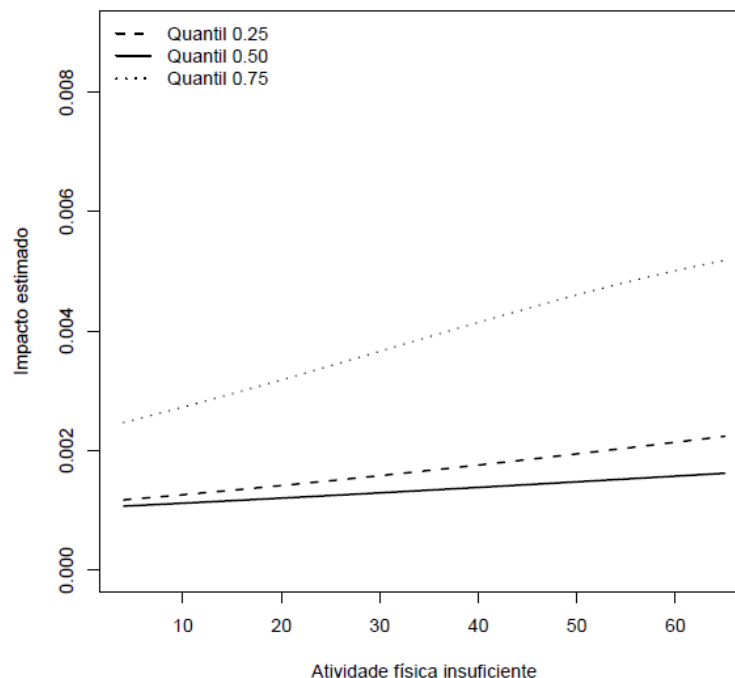
Com o objetivo de estimar as curvas de impacto ou efeito marginal da atividade física insuficiente sobre a proporção de adultos obesos nas nações considerou-se três situações, como apresentado na Figura 3.

Ou seja, os ajustes referentes aos quantis 0,25, 0,50 e 0,75. Além disso as covariáveis URB e ALC estão fixadas na mediana.

Dessa forma, é possível variar os valores de INAT para determinar o aumento

provocado na variável resposta. Como resultado, observa-se que o impacto é positivo e cresce lentamente à medida que se aumenta os valores de INAT.

A curva estimada no quantil 0,25 apresenta um impacto maior que no quantil 0,50, enquanto o maior impacto é observado no quantil 0,75. Por fim, é observado que não existem grandes diferenças entre as curvas nos quantis 0,25 e 0,50.



**Figura 3** - Efeito marginal da atividade física insuficiente sobre a proporção de adultos obesos considerando os ajustes nos quantis 0,25, 0,50 e 0,75.

## CONCLUSÃO

Neste artigo, foi utilizada uma base de dados referente à obesidade adulta em 78 nações em 2014.

A análise inicial foi realizada por meio da construção de mapas que possibilitaram visualizar a distribuição dos valores das covariáveis abordadas no estudo.

Como resultado verificou-se que os continentes Europa, América e Oceania apresentaram altos valores para a proporção de adultos obesos. Em particular, os Estados Unidos, a República Checa e a Nova Zelândia apresentaram os maiores valores entre esses três continentes.

A partir da modelagem da obesidade adulta nas nações, utilizando o modelo de regressão quantílica, verificou-se que as covariáveis porcentagem de atividade física insuficiente, consumo médio de álcool por pessoa em um ano e porcentagem da população que vivem em áreas urbanas tendem a aumentar as proporções de adultos obesos nas nações.

Adicionalmente, ao ajustar os modelos para diferentes quantis observou-se, por meio do pseudo- $R^2$ , que em alguns pontos da distribuição condicional da variável resposta a contribuição destas covariáveis para explicar o desfecho não é a mesma.

Por fim, estimou-se o impacto da atividade física insuficiente em diferentes cenários e constatou-se que o impacto é positivo e cresce lentamente à medida que se aumenta os valores desta covariável.

Além disso, verifica-se que com o modelo de regressão quantílica os gestores possuem mais uma alternativa para tomada de decisão na realização de medidas de controle/prevenção da doença.

## REFERÊNCIAS

- 1-Amaral, O.; Pereira, C. Obesidade da genética ao ambiente. *Millenium*. Num. 34. 2008. p.311-22.
- 2-Anjos, L. Obesidade e saúde pública. *Fiocruz*. 2006.
- 3-Aranceta-Bartrina, J.; Pérez-Rodrigo C.; Alberdi-Aresti, G.; Ramos-Carrera, N.; Lázaro-4-Bottai, M.; Cai, B.; Mckeown, R. Logistic quantile for bounded outcomes. *Statistics in Medicine*. Vol.29. Num.2. 2010. p.309-17.

5-Buchinsky, M. Recent advances in quantile regression models: a practical guideline for empirical research. *The Journal of Human Resources*. Vol.33. Num.1. 1998. p.88-126.

6-Callo, G.; Gigante, D.P.; Barros, F.C.; Horta, B.L. Excesso de peso/obesidade no ciclo da vida e composição corporal na idade adulta: coorte de nascimentos de Pelotas, Rio Grande do Sul, Brasil, 1982. *Caderno de Saúde Pública*. Vol.32. Num.4. 2016. p.e00174014.

7-Carvalho, A.R.M.; Belém, M.O.; Oda, J.Y. Sobrepeso e obesidade em alunos de 6-10 anos de escola Estadual de Umuarama/ PR. *Arquivos de Ciências da Saúde UNIPAR*. Vol.21. Num.1. 2017. p.3-12.

8-Cordeiro, J.P.; Dalmaso, S.B.; Anceschi, S.A.; Sá, F.G.S.; Ferreira, L.G.; Cunha, C.M.R.H.; Leopoldo, A.S.; Lima-Leopoldo, A.P. Hipertensão em estudantes da rede pública de Vitória/ES: influência do sobrepeso e obesidade. *Revista Brasileira de Medicina do Esporte*. Vol.22. Num.1. 2016. p.59-65.

9-Feizi, A.; Aliyari, R.; Roohafza, H. Association of perceived stress with stressful life events, lifestyle and sociodemographic factors: a large-scale community-based study using logistic quantile regression. *Computational and Mathematical Methods in Medicine*. Vol.2012. 2012. p.1-12.

10-Garcia, B.C.C.; Matos, M.A.; Pinhal, K.C.; Oliveira, M.F.A.; Almeida, R.T.; Bento, M.E.; Fernandes, V.O.; Magalhães, F.C.; Esteves, E.A.; Rocha-Vieira, E.; Amorim, F.T. A obesidade na população adulta de Diamantina: orientações para a saúde, caracterização antropométrica, bioquímica e de aptidão cardiorrespiratória. *Revista Brasileira de Extensão Universitária*. Vol. 7. Num. 2. 2016. p. 129-137.

11-Geraci, M. Qtools: a useful package for quantiles. 2016. Disponível em: <<https://cran.r-project.org/web/packages/Qtools/vignettes/Qtools.pdf>>. Acesso em: 21/11/2018.

12-Hao, L.; Naiman, D. Quantile regression. Sage publications. 2007.

13-He, X.; Zhu, L. A lack-of-fit test for quantile regression. *Journal of the American Statistical Association*. Vol.98. Num.464. 2003. p.1013-22.

- 14-Koenker, R. Quantile regression. Cambridge University Press. 2005.
- 15-Koenker, R.; Bassett, G. Regression quantiles. *Econometrica*. Vol.46. Num.1. 1978. p.33-50.
- 16-Koenker, R. Censored quantile regression redux. *Journal of Statistical Software*. Vol.27. Num.6. 2008. p.1-26.
- 17-Koenker, R.; Machado, J. Goodness of and related inference processes for quantile regression. *Journal of the American Statistical Association*. Vol.94. Num.448. 1999. p.1296-310.
- 18-Koenker, R.; Bassett, G. Robust tests for heteroscedasticity based on regression quantiles. *Econometrica*. Vol.50. Num.1. 1982. p.43-61.
- 19-Lima, N.P.; Horta, B.L.; Motta, J.V.S.; Valença, M.S.; Oliveira, V.; Santos, T.V.; Gigante, D.P.; Barros, F.C. Evolução do excesso de peso e obesidade até a idade adulta, Pelotas, Rio Grande do Sul, Brasil, 1982-2012. *Caderno de Saúde Pública*. Vol.31. Num.9. 2015. p.2017-2025.
- 20-Linhares, R.S.; Horta, B.L.; Gigante, D.P.; Dias-da-Costa, J.S.; Olinto, M.T.A. Distribuição de obesidade geral e abdominal em adultos de uma cidade no Sul do Brasil. *Caderno de Saúde Pública*. Vol.28. Num.3. 2012. p.438-447.
- 21-Luca, F.D.; Boccuzzo, G. What do healthcareworkers know about sudden infant death syndrome?: The results of the Italian campaign 'Genitoripiù'. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*. Vol.177. Num.1. 2014. p.63-82.
- 22-Mccullagh, P.; Nelder, J. Generalized linear models. London. Chapman and Hall. 1989.
- 23-Mua, Y.; Hea, X. Power transformation toward a linear regression quantile. *Journal of the American Statistical Association*. Vol.102. Num.477. 2007. p.37-41.
- 24-Neyman, J.; Pearson, E. On the use and interpretation of certain test criteria for purposes of statistical inference. *Biometrika*. Vol.20. 1928. p.175-240.
- 25-Ng, M.; Fleming, T.; Robinson, M.; Thomson, B.; Graetz, N.; Margono, C.; Mullany, E.C.; Biryukov, S.; Abbafati, C.; Abera, S.F.; Abraham, J.P.; Abu-Rmeileh, N.M.; Achoki, T.; AlBuhairan, F.S.; Alemu, Z.A.; Alfonso, R.; Ali, M.K.; Ali, R.; Guzman, N.A.; Ammar, W.; Anvari, P.; Banerjee, A.; Barquera, S.; Basu, S.; Bennett, D.A.; Bhutta, Z.; Blore, J.; Cabral, N.; Nonato, I.C.; Chang, J.C.; Chowdhury, R.; Courville, K.J.; Criqui, M.H.; Cundiff, D.K.; Dabhadkar, K.C.; Dandona, L.; Davis, A.; Dayama, A.; Dharmaratne, S.D.; Ding, E.L.; Durrani, A.M.; Esteghamati, A.; Farzadfar, F.; Fay, D.F.; Feigin, V.L.; Flaxman, A.; Forouzanfar, M.H.; Goto, A.; Green, M.A.; Gupta, R.; Hafezi-Nejad, N.; Hankey, G.J.; Harewood, H.C.; Havmoeller, R.; Hay, S.; Hernandez, L.; Hussein, A.; Idrisov, B.T.; Ikeda, N.; Islami, F.; Jahangir, E.; Jassal, S.K.; Jee, S.H.; Jeffreys, M.; Jonas, J.B.; Kabagambe, E.K.; Khalifa, S.E.; Kengne, A.P.; Khader, Y.S.; Khang, Y.H.; Kim, D.; Kimokoti, R.W.; Kinge, J.M.; Kokubo, Y.; Kosen, S.; Kwan, G.; Lai, T.; Leinsalu, M.; Li, Y.; Liang, X.; Liu, S.; Logroscino, G.; Lotufo, P.A.; Lu, Y.; Ma, J.; Mainoo, N.K.; Mensah, G.A.; Merriman, T.R.; Mokdad, A.H.; Moschandreas, J.; Naghavi, M.; Naheed, A.; Nand, D.; Narayan, K.M.; Nelson, E.L.; Neuhouser, M.L.; Nisar, M.I.; Ohkubo, T.; Oti, S.O.; Pedroza, A.; Prabhakaran, D.; Roy, N.; Sampson, U.; Seo, H.; Sepanlou, S.G.; Shibuya, K.; Shiri, R.; Shiue, I.; Singh, G.M.; Singh, J.A.; Skirbekk, V.; Stapelberg, N.J.; Sturua, L.; Sykes, B.L.; Tobias, M.; Tran, B.X.; Trasande, L.; Toyoshima, H.; Van de Vijver, S.; Vasankari, T.J.; Veerman, J.L.; Velasquez-Melendez, G.; Vlassov, V.V.; Vollset, S.E.; Vos, T.; Wang, C.; Wang, X.; Weiderpass, E.; Werdecker, A.; Wright, J.L.; Yang, Y.C.; Yatsuya, H.; Yoon, J.; Yoon, S.J.; Zhao, Y.; Zhou, M.; Zhu, S.; Lopez, A.D.; Murray, C.J.; Gakidou, E. Global, regional and national prevalence of overweight and obesity in children and adults 1980-2013: A systematic analysis. *Lancet*. Vol.384. Num.9945. 2014. p.766-81.
- 26-Pereira, M.S.; Schaefer, J.C.F.; Dutra, M.C.; Uliano, E.J.M.; Jesus, C.R.; Silvestre, B.C.; Machado, D.F.G.P.; Trevisol, D.J.; Schuelter-Trevisol, F. Prevalência de obesidade geral e central na população adulta na cidade de Tubarão, Santa Catarina: um estudo de base populacional. *Scientia Medica*. Vol.24. Num.3. 2014. p.245-53.

27-Powell, J. Censored regression quantiles. Journal of Econometrics. Vol.32. 1986. p.143-55.

28-R Core Team. R: A language and Environment for Statistical Computing. Vienna, Austria. 2013. Disponível em: <<http://www.R-project.org/>>. Acesso em: 21/11/2018.

29-Salas, M.M.; Gómez, L.R.; Martínez, D.E.; Soto, J.J.J.; Martín-Ramiro, J.J. Relacion entre la jornada laboral y las horas de sueño com el sobrepeso y la obesidade em la población adulta española según los datos de la Encuesta Nacional de Salud 2012. Revista Espanhola de Salud Pública. Vol.91. 2017. p.1-10.

30-Santos, B. Modelos de regressão quantílica. Dissertação de Mestrado. São Paulo. 2012.

31-Souza, S. A. Modelagem da obesidade adulta nas nações: uma análise via modelos de regressão beta e quantílica. Dissertação de Mestrado. Universidade Federal da Paraíba. João Pessoa. 2017.

32-Zapata, M.E.; Bibiloni M.M.; Tur, J.A. Prevalence of overweight, obesity, abdominal-obesity and short stature of adult population of Rosario, Argentina. Nutrición Hospitalaria. Vol.33. Num.5. 2016. p.1149-1158.

#### **CONFLITO DE INTERESSE**

Não há conflito de interesse.

Recebido para publicação em 22/05/2019

Aceito em 21/06/2019