

UNIVERSIDADE ESTADUAL DE MARINGÁ  
DEPARTAMENTO DE ESTATÍSTICA  
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOESTATÍSTICA

EMERSON BARILI

MODELOS PARA ACOMODAR A  
SOBREDISPERSÃO E IDENTIFICAR  
ÍNDICES NUTRICIONAIS PARA POPULAÇÃO  
INDÍGENA

Maringá - Paraná  
2015.

EMERSON BARILI

MODELOS PARA ACOMODAR A SOBREDISPERSÃO E  
IDENTIFICAR ÍNDICES NUTRICIONAIS PARA  
POPULAÇÃO INDÍGENA

Dissertação apresentada ao Programa de  
Pós-Graduação em Bioestatística como  
requisito para Obtenção do Título de  
Mestre em Bioestatística.

Orientadora: Prof.<sup>a</sup> Dra. Rosângela  
Getirana Santana.

Coorientadora: Prof.<sup>a</sup> Dra. Taqueco  
Teruya Uchimura.

Maringá - Paraná  
2015.

EMERSON BARILI

MODELOS PARA ACOMODAR A SOBREDISPERSÃO E  
IDENTIFICAR ÍNDICES NUTRICIONAIS PARA  
POPULAÇÃO INDÍGENA

Dissertação apresentada ao Programa de  
Pós-Graduação em Bioestatística como  
requisito para Obtenção do Título de  
Mestre em Bioestatística.

Orientadora: Prof.<sup>a</sup> Dra. Rosângela  
Getirana Santana.

Coorientadora: Prof.<sup>a</sup> Dra. Taqueco  
Teruya Uchimura.

Trabalho aprovado.  
Maringá - Paraná, Setembro de 2015:

---

Profa. Dra. Rosângela Getirana Santana  
Orientadora

---

Profa. Dra. Taqueco Teruya Uchimura  
Coorientadora

---

Prof. Dr. Silvano César da Costa  
Membro 2

---

Profa. Dra. Izolde Terezinha S. Previdelli  
Membro 3

Maringá - Paraná  
2015.

## Apresentação

Sou matemático, graduado pela Universidade Estadual de Maringá, (UEM) em 2007. Fui professor nos Colégios Estadual Márcia Vaz Tostes de Abreu e Urbano Pedroni por 5 anos lecionando as disciplinas de Física e Matemática para alunos das séries do Ensino Médio e professor colaborador do Departamento de Matemática da Universidade Estadual de Maringá. Neste período me identifiquei com matemática aplicada observando a forte necessidade de uma proposta de estudo que transmita ao aluno a real importância da Matemática e da Estatística no mundo atual e resolvi fazer o Mestrado em Bioestatística para ampliar o conhecimento na área da estatística podendo contribuir nas pesquisas principalmente no âmbito da saúde.

Este estudo foi desenvolvido com duas propostas distintas com diferentes pretensões, apresentadas em dois projetos. O primeiro projeto sob a orientação da Professora Dra. Rosângela Getirana Santana se refere a metodologia estatística adotada para tratar da sobredispersão, fenômeno presente em dados de contagem que pode acarretar em ajustes de modelos imprecisos. Para investigar esta metodologia utilizou-se os dados da pesquisa Avaliação da Suplementação com Quercetina sobre as células Imunorreativas da Serotonina e Neurônios Entéricos dos Segmentos Jejunal de ratos Diabéticos induzidos por Estreptozotocina, coordenada pela Prof. Juliana Colombo do Departamento de Farmácia da Universidade Estadual de Maringá.

O segundo projeto, mais envolvido com a esfera da saúde, a pesquisa foi realizada em dois estágios, ou seja, sendo a primeira parte dedicada para verificar as condições ambientais e sócio-econômicas da comunidade indígena Kaingang do Ivaí e a coleta dos dados, e a segunda parte buscou-se identificar os índices antropométricos para avaliar o estado nutricional desta população indígena, através de uma estatística que não possui relação com a estatística abordada no capítulo anterior. Este capítulo foi orientado pela Professora Dra. Taquaco Teruya Uchimura. Minha participação neste projeto se pontuou em todas as etapas, iniciando-se pelo planejamento das atividades a serem desenvolvidas na Terra Indígena Kaingang, coleta dos dados antropométricos da população no local, registros dos dados e posteriormente a elaboração do banco destes dados, realização das análises para melhor apresentar os resultados e divulgar os dados em revistas especializadas.

Este trabalho é todo dedicado a minha esposa Cirlene e a minha filha Alícia.

“Não consigo conceber um deus eu recompense e puna as suas criaturas, nem que tenha uma vontade do tipo que experimentamos em nós mesmos. Não consigo, nem quero conceber um indivíduo que sobreviva a sua morte física; que as almas fracas, por medo ou egoísmo absurdo, alimentem esses pensamentos. Eu me satisfaço com o mistério da eternidade da vida e com o vislumbre da maravilhosa estrutura do mundo real, junto com o esforço diligente de compreender uma parte, por menor que seja, da Razão que se manifesta na natureza”.

*Albert Einstein*

“Não há linguagem mais universal e mais simples, mais livre de erros e de obscuridade, isto é, mais digna de expressar as relações invariáveis das coisas naturais [...] [A matemática] parece ser uma faculdade da mente humana destinada a suplementar a brevidade da vida e a imperfeição dos sentidos.”

Joseph Fourier

## Agradecimentos

Nesta oportunidade agradeço a Deus por todas as bênçãos recebidas, por me mostrar que sou protegido, guiado e iluminado pela sua presença no mais íntimo do meu ser. À minha mãe “Luzia”, minha avó “Luzinete” e ao meu tio “José Serafim” por toda as lutas, grandezas, dedicações e amor para o meu crescimento humano, intelectual e espiritual.

À minha esposa “Cirlene” pelo carinho, amor, paciência e apoio que dera nos momentos difíceis ao longo da minha vida.

Ao Programa de Pós-Graduação em Bioestatística da Universidade Estadual de Maringá -UEM, pelo apoio à minha participação no mestrado.

À todos os professores do Departamento de Estatística da Universidade Estadual de Maringá que contribuíram para a minha formação.

Um agradecimento especial às minhas orientadoras e amigas Professoras “Rosângela Getirana” e “Taqueto Teruya Uchimura” pelos apoios amizados, além de vossas dedicações, competências e especiais atenção nas revisões e sugestões, fatores importantes para conclusão deste trabalho.

À Professora “Isolde Previdelli” pela amizade, apoio, compreensão, paciência e sugestões relevantes no desenvolvimento desse estudo.

Aos amigos “Omar”, “Paulo”, “Rafaela”, “Marina”, “Matheus”, “Guilherme”, “ Marcos” e todos os outros pelo companheirismo ao longo dessa jornada.

Aos velhos amigos e parceiros “Sérgio”, “Eduardo”, “Emílio” pelos muitos anos de amizade, carinho e apoio diante dos desafios.

À aluna “Adriana Mayumi Moriwaki” pela sua determinada e eficiente participação na realização deste estudo.

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico - CNPq, pelo incentivo financeiro.

# Lista de Figuras

## **Projeto 1: Acomodando a sobredispersão com formas mais gerais para a função de variância.**

Figura 1 – Médias e erros-padrão do número de neurônios por grupos e classes. . . .	21.
Figura 2 - Distância de Cook . . . . .	24.
Figura 3 – Envelopes dos três modelos. . . . .	24.

## **Projeto 2: Indicadores de referência para a avaliação nutricional específicos para população Indígena Kaingang-Brasil, 2014.**

Figura 1 - Curvas de sensibilidade e especificidade (TG-ROC) para pontos de corte. . . . .	.39.
--	------



## Lista de Tabelas

### **Projeto 1 : Acomodando a sobredispersão com formas mais gerais para a função de variância.**

Tabela 1 - Estimativas, Erros-Padrão e P-valores para os três modelos. . . . .	.22
Tabela 2 - Critério de informação de ajuste modelo de Poisson ( <i>proc</i> GENMOD . .	23
Tabela 3 - Critério de informação de ajuste modelos Quase-Poisson e Binomial Negativa. . . . .	23
Tabela 4 - Análise do Tipo III de efeitos fixos. . . . .	.25
Tabela 5 - Critérios para avaliação dos modelos. . . . .	.25

### **Projeto 2: Indicadores de referência para a avaliação nutricional específicos para a população Indígena Kaingang-Brasil, 2014.**

Tabela 1 - Adistribuição da população segundo a faixa etária, sexo e idades máxima, mínima e média $\pm$ DP. . . . .	38
Tabela 2 - Calibração dos índices antropométricos (percentil e <i>escore-z</i> ). . . . .	40
Tabela 3 - Distribuição da população nas faixa etária, segundo classificação dos índices. . . . .	41
Tabela 4 - Classificação da população segundo os índices antropométricos, valores críticos do SISVAN, K-means e p-valor. . . . .	.42
Tabela 5 - Distribuição da população com valores dos testes diagnósticos. . . . .	43

# Sumário

## **Projeto 1. Acomodando a sobredispersão com formas mais gerais para a função de variância.**

1. Introdução .....	13
2. Metodologia .....	15
2.1 Seleção de Modelos .....	18
3. Resultados .....	20
4. Discussão .....	25
5. Apêndice .....	27
6. Referências .....	28

## **Projeto 2. Indicadores de referência para a avaliação nutricional específicos para a população Indígena Kaingang-Brasil, 2014.**

1. Introdução .....	33
2. Metodologia .....	35
3. Resultados .....	38
4. Discussão .....	44
5. Conclusão .....	48
6. Apêndice .....	49
7. Referências .....	51
8. Termo de consentimento livre e esclarecimento .....	55
9. Parecer substanciado do CEP .....	57

## Projeto 1: Acomodando a sobredispersão com formas mais gerais para a função de variância

### Resumo

Neste estudo a variável resposta de interesse é a contagem que possui variância maior que a variância da distribuição de Poisson dos modelos lineares generalizados sugerida, ou seja, os dados apresentaram sobredispersão. Para acomodar esse fenômeno foi proposta uma função de variância que não possui distribuição de probabilidade conhecida na família exponencial, e nesse caso, utilizou-se o método de quase-verossimilhança, que estima o parâmetro de dispersão pelo método dos momentos que não fornecem os erros-padrão das estimativas, portanto, os valores destes erros foram obtidos usando o método de pseudo-verossimilhança considerando apenas a parte do efeito fixo, isto é, os erros-padrão das estimativas ajustando os parâmetros da matriz de estrutura de covariância  $\mathbf{R}$  dos efeitos fixos. Outra função de variância proposta é da distribuição de probabilidade Binomial Negativa obtendo as estimativas dos parâmetros pelo método de máxima verossimilhança. A análise foi realizada usando as *procedures* GENMOD e GLIMMIX do *software* SAS e o ajuste dos modelos foram avaliados pela análise de resíduos e por testes estatísticos como o teste F e o teste Qui-quadrado.

**Palavras-chaves:** Sobredispersão, pseudo-verossimilhança, funções de variância, modelos lineares generalizados, matriz de covariância.

## Abstract

In this study the response variable of interest is the count which has a greater variance than the Poisson distribution the variance of generalized linear models suggested, that is, the data presented overdispersion. To accommodate this phenomenon was proposed a variance function that has no known probability distribution in the exponential family, and in this case, used the method of quasi-likelihood estimating dispersion parameter by the method of moments that do not provide the error-standard estimates, therefore, the values of these errors were obtained using pseudo-likelihood method, considering only that part of the fixed end, that is, standard errors of the estimate by adjusting the parameters of the structure R covariance matrix of fixed effects. Another variance function proposed is negative binomial probability distribution obtaining estimates of parameters by maximum likelihood method. The analysis was performed using the procedures GENMOD and GLIMMIX of SAS software and the adjustment of the models were evaluated by statistical analysis of residues and tests such as the F-test and the chi-square test.

**Keywords:** Overdispersion, pseudo-likelihood, variance functions, generalized linear models, covariance matrix.

## 1. Introdução

O modelo linear generalizado com componente aleatório Poisson é comumente utilizado para análise de dados com variável resposta obtida por resultado de contagem. Nesta pesquisa a variância nominal da variável resposta é maior que a predita pela distribuição de probabilidade assumida Poisson, e quando isso ocorre diz-se que os dados apresentam sobredispersão, que também pode ser observada quando a *deviance* (D) é maior que os graus de liberdade (GL), isto é,  $D/GL > 1$  Hinde e Demétrio (1998).

Possíveis causas para a sobredispersão foram descritas por Hinde e Demétrio (1998) como a variabilidade do material experimental, correlação entre indivíduos, delineamento amostral por "*cluster*", excessos de zero ou omissão de variáveis preditoras. Na literatura, as causas desse fenômeno foram destacadas por vários pesquisadores, entre eles Lester (2012) e Bonafede et al. (2014) que notaram que a sobredispersão foi proveniente das medidas serem repetidas Mohebbi, Wolfe e Forbes (2014) observaram que a da sobredispersão foi devido a alta correlação entre as variáveis e Alam (2015), verificou que a sobredispersão foi em virtude do excesso de zeros. Nos modelos em que a sobredispersão não foi acomodada, os erros-padrão geralmente são subestimados o que pode causar interpretações errôneas dos parâmetros do preditor linear, tendo como consequência as tomadas de decisões incorretas.

Para acomodar a sobredispersão várias metodologias foram propostas e, entre elas, a apresentada por McCullagh e Nelder (1989) que propuseram o método de assumir uma forma mais geral para a função de variância, possivelmente incluindo parâmetros adicionais, enquanto Hinde e Demétrio (1998) sugerem modelos em dois estágios, isto é, supondo que o próprio parâmetro básico tem alguma distribuição. Lindén e Mäntyniemi (2011) sugeriram uma parametrização da distribuição Binomial Negativa em que dois parâmetros sobredispersos são introduzidos permitindo a relação quadrática entre a média e a variância, Ridout, Hinde e Demétrio (2001), acomodou a sobredispersão gerada pelo excesso de zeros, utilizando a distribuição Binomial Negativa Hurdle obtendo valores dos parâmetros ajustado pelo método de máxima verossimilhança. Por outro lado, Bonafede et al. (2014) acomodaram a sobredispersão

pelo método de máxima verossimilhança utilizando a distribuição Binomial Negativa. Zhang et al. (2015) em estudo sobre o número de ação e reação da polimerase digital acomodaram a sobredispersão usando métodos de medidas repetidas de uma forma não paramétrica. Kassahun et al.(2014) apontaram que, em uma abordagem hierárquica, a sobredispersão pode ser acomodada por modelos Poisson com efeitos aleatórios seguindo distribuição Gama. Hoef e Boveng (2007) adotaram duas funções de variância para estimar os parâmetros de dispersão.

Este estudo objetivou propor funções de variância com parâmetros de escala e dispersão para acomodar a sobredispersão apresentada pelos dados, cuja variável resposta de interesse é o número de neurônios do intestino de ratos utilizados em um experimento que investiga o efeito de flavonóides (grupo de substâncias químicas presente em diversos vegetais que fazem parte da alimentação humana. Cebola e uva) no tratamento de “diabetes mellitus”.

Dessa forma, para obter as estimativas dos parâmetros utilizou-se métodos como o de máxima verossimilhança, método dos momentos, método de quasi-verossimilhança utilizados nos trabalhos de Wooldridge (2014) em modelos não lineares e de Yu, Bai e Ding (2015) que aplicaram este método em modelos mistos, e ainda optou-se pelo método de estimação pseudo-verossimilhança apresentado por Besag (1974) que definiu a técnica como o produto de distribuições condicionais para cada variável dada todas as outras. Para esse método utilizou-se a *procedure* GLIMMIX do *software* SAS 9.3. Ressaltando que os dados utilizados nesta pesquisa não possui efeito aleatório. O método pseudo-verossimilhança pode ser verificado nos trabalhos como de Seemayer, Gruber e Söding (2014) que utilizaram o método para separar pares de variáveis correlacionadas, Qiu, Nordman e Vardeman (2014) que introduziram o método de pseudo-verossimilhança para tratar de dados incompletos, Varin, Reid e Firth (2011) que demonstraram que as estimativas dos parâmetros são consistentes e assintoticamente seguem distribuições normais. Chen, Hong e Shum (2007) que mostraram que o método pseudo-verossimilhança fornece estimativas imparciais, erros padrão bem estimados e intervalos de confiança com alta probabilidade de abrangência.

Um dos testes amplamente adotado para a seleção de modelos é o teste da razão de verossimilhança, utilizado para seleção de modelos encaixados, isto é, modelos em que um pode ser obtido a partir do outro impondo-se um conjunto de restrições lineares no vetor de parâmetros. Por outro lado, tem-se os modelos não encaixados, que são aqueles que não podem ser obtidos a partir de outro impondo um conjunto de restrições lineares do vetor de parâmetros. Além dos testes é

também comum o uso de critérios para selecionar modelos como o Critério de Informação de Akaike (AIC) Bozdogan (1987) e o Critério de Informação Bayesiano (BIC) Schwarz et al.(1978) que apresentam característica semelhante ao AIC, porém, é penalizado pelo número de parâmetros e pelo tamanho da amostra. Apesar de serem comumente utilizados, é importante ter cautela quanto ao uso destes critérios para modelos não encaixados, ou seja, não utilizá-los com muito rigor, porém, podem contribuir com o processo de construção do modelo. No entanto, para os modelos concorrentes não encaixados é comum o uso do teste de Vuong (1989) que necessita das funções de log-verossimilhança dos modelos. Neste estudo porém, não se adotou este modelo, pois, não se pode obter a função de log-verossimilhança quando não se conhece a distribuição.

No contexto da análise de ajuste de modelos, verifica-se a análise de resíduos, testes estatísticos como o teste da razão de verossimilhança que segue a distribuição F-Snedecor, realiza-se análise do Tipo 3 que fornece os estimadores pelo método de máxima verossimilhança avaliando o contraste entre os modelos, identifica os *outliers* que podem afetar significativamente o modelo obtidos pelas estatísticas *leverage*, distância de Cook e DFITS. Nesta pesquisa o diagnóstico foi obtido utilizando as *procedures* GENMOD e GLIMMIX do software SAS 9.3.

## 2. Metodologia

Neste estudo, os dados foram analisados tendo como variável resposta  $Y$  a

contagem. Assim, utilizou-se a distribuição de probabilidade Poisson com esperança  $E(Y) = \mu$  e função de variância  $V(Y) = \mu$ , sendo as estimativas obtidas pelo método de máxima verossimilhança. Geralmente, o modelo Poisson não se ajusta bem aos dados na presença de sobredispersão, Desta forma, para acomodar esse fenômeno foram propostas as funções de variância dada por

$$V(Y) = \phi\mu \quad \text{e} \quad V(Y) = \mu(1 + \kappa\mu)$$

sendo  $\mu$  a média,  $\phi$  e  $\kappa$  os parâmetros de dispersão. Nesta pesquisa a componente sistemática é a mesma para todos os modelos que, segundo Demétrio (2002) possui a forma de uma soma linear dada por

$$\eta_i = g(\mu_i) = \sum x_{ij}\beta_j = x_i\beta$$

Com  $i = 1, \dots, n$ , sendo  $n = (n_1, n_2, \dots, n_n)^T$  o preditor linear,  $g(\cdot)$  a função de ligação que no caso do modelo Poisson, é a logarítmica:  $g(\mu_i) = \log(\mu_i)$ . O vetor  $x_i = (x_{i1}, x_{i2}, \dots, x_{ip})^T$  representa as variáveis preditoras e  $\beta = (x_1, x_2, \dots, x_p)^T$  o vetor dos parâmetros de regressão, estimados pelo método de máxima verossimilhança.

No entanto, nos modelos lineares generalizados, a função de variância  $V(Y) = \phi\mu$  não possui distribuição de probabilidade conhecida, com isso, adotou-se o método de quase-verossimilhança, que considera o parâmetro de dispersão  $\phi$  como uma constante, assim, este método não permite acomodar efeitos associados a este parâmetro, porém influencia na demanda de outro método para obter o valor estimado de  $\phi$  (McCullagh & Nelder, 1989). Portanto, o logaritmo da função de quase-verossimilhança  $Q$  é definida como

$$Q(\mu_i, y_i) = \frac{1}{\phi} \sum_{i=1}^n \int_y^{\mu_i} \frac{(y_i - \mu_i)}{V(\mu_i)} d\mu_i$$

em que  $y_i$  é o valor de cada observação e  $\mu_i$  representa a média. Utilizando esse método com a função de variância  $V(Y) = \phi\mu$ , pôde-se observar que o *kernel* do logaritmo da função quase-verossimilhança é igual a função de log-verossimilhança da distribuição Poisson. Desta forma, os parâmetros de regressão foram estimados pelo método de máxima verossimilhança e o valor estimado de  $\phi$  foi obtido pelo método dos momentos definido por

$$\hat{\phi} = \frac{1}{n-p} \sum_{i=1}^n \frac{(Y_i - \hat{\mu}_i)^2}{V(\hat{\mu}_i)}$$



sendo  $\mu$  a estimativa do valor esperado de  $Y$ . A eficiência do método de quase-verossimilhança é verificada pelo critério de *deviance*, que para este método denomina-se *quase-deviance* dada pela seguinte expressão (PAULA, 2004)

$$\begin{aligned} D(y, \hat{\mu}) &= 2\phi \{Q(y, y) - Q(\hat{\mu}, y)\} \\ &= -2\phi Q(\hat{\mu}, y) \\ &= -2\phi \sum_{i=1}^n Q(\hat{\mu}_i, y_i) \\ &= 2\phi \sum_{i=1}^n \int_{\hat{\mu}_i}^{y_i} \frac{y_i - t}{V(t)} dt \end{aligned}$$

que não depende de  $\phi$  sendo  $Q(y, y)$  e  $Q(\hat{\mu}, y)$  funções de quase-verossimilhança do modelo completo e do modelo em investigação, respectivamente. Embora não seja conhecida a distribuição nula para  $D(y, \hat{\mu})$  é comum compará-la com os percentis da distribuição  $\chi_{n-p}^2$ , em que  $(n-p)$  são os graus de liberdade (PAULA, 2004). Além disso, possui a eficiência de comparar modelos com diferentes preditores lineares ou funções de ligação.

Usando *procedure* GENMOD obteve-se o valor do parâmetro de escala  $\phi$  pelo método dos momentos. Outra forma usada para obter a estimativa desse parâmetro foi usar a técnica de estimação de máxima pseudo-verossimilhança implementada na *procedure* GLIMMIX do *software* SAS 9.3 verificada por Morel e Neerchal (2012), que apresentaram resultados considerando apenas a matriz dos efeitos fixos. Dessa forma, se o vetor de parâmetros  $\gamma$  da parte aleatórios modelos lineares generalizados mistos é igual a zero e  $R$  (matriz de covariância) é definida como  $R = \phi I$ , o modelo se reduz a um modelo linear generalizado. Se uma distribuição de probabilidade condicional contém o parâmetro de escala  $\phi \neq 1$ , a variância do método pseudo-verossimilhança é dada por

$$V(\theta^*) = \Delta^{-1} A^{\frac{1}{2}} R^* A^{\frac{1}{2}} \Delta^{-1}$$

sendo  $\theta^*$  o vetor de parâmetros da covariância com  $q-1$  elementos,  $\Delta = \left( \frac{\partial g^{-1}(\eta)}{\partial \eta} \right)_{\beta, \gamma}$  e  $R^*$  a matriz reparametrizada de  $R$ . Para a

parametrização de  $R = \phi I$ , então  $\theta^*$  contém a razão da componente da variância e o parâmetro  $\phi$  e  $R^* = I$ . O parâmetro  $\phi$  é estimado pela equação

$$\hat{\phi} = \hat{r} \hat{V}(\theta^*) - 1\hat{r} / m$$

em que  $m$  é a soma da frequência usada na análise,  $\hat{r}$  é o valor estimado de

$$r = p - (XV^{-1}X)^{-1} XV^{-1}p \text{ em que } \hat{p} \text{ são valores estimados do pseudo modelo } P = X\beta + \epsilon$$

(WOLFINGER; TOBIAS; SALL, 1994).

A outra função que foi proposta para acomodar a sobredispersão é a função de variância da distribuição Binomial Negativa, que estimou os parâmetros pelo método de máxima verossimilhança através da função de log-verossimilhança parametrizada na *procedure* GENMOD conforme a função

$$l(y_i, \mu_i, \kappa) = y_i \log \left[ \frac{\kappa \mu_i}{w_i} \right] \left( \frac{\kappa \mu_i}{w_i} \right) - (y_i + w_i / \kappa) \log \left( 1 + \frac{\kappa \mu_i}{w_i} \right) + \log \left( \frac{\Gamma(y_i + w_i / \kappa)}{\Gamma(w_i / \kappa)} \right)$$

sendo  $\mu$  a média,  $w_i$  o peso para cada observação, que nesse estudo  $w_i = 1$ , e  $\kappa$  o parâmetro de dispersão. Para essa parametrização da função log-verossimilhança expressa como os valores preditos da média e do vetor da variável resposta, a adequação do modelo é analisada pelo critério de *deviance* definida por

$$D(y, \mu) = 2 \sum_{i=1}^n \left[ y \log \left( \frac{y}{\mu} \right) - \left( y + \frac{w_i}{\kappa} \right) \log \left( \frac{y + \frac{w_i}{\kappa}}{\mu + \frac{w_i}{\kappa}} \right) \right].$$

## 2.1 Seleção de Modelos

Dentre os critérios para a seleção de modelos que se baseiam na função de máxima verossimilhança utilizados em modelos encaixados, estão o teste da razão de verossimilhança (TRV), critério de informação de Akaike (AIC), critério Bayesiano de Schwarz (BIC). Se os modelos não são encaixados o teste da razão de

verossimilhança não segue uma distribuição  $\chi^2_{n-p}$  assintoticamente, o que geralmente acontece no caso dos modelos encaixados. O teste da razão de verossimilhança necessita da estimação do modelo restrito e do modelo sem restrição. Considere  $\Theta$  um espaço paramétrico, e um vetor pertencente a  $\Theta$  denotado por

$$\theta = (\theta_1, \theta_2, \dots, \theta_k)$$

sendo  $\theta$  um vetor de parâmetros do modelo saturado e  $\hat{\theta}$  o vetor de parâmetros do modelo restrito, sendo suas estimativas, obtidas pelo método de máxima verossimilhança  $L(\theta)$  e  $L(\hat{\theta})$  respectivamente e para  $H_0$  verdadeiro tem-se o teste dados por

$$LR = -2 \left( \log(L(\theta)) - \log(L(\hat{\theta})) \right),$$

que assintoticamente segue uma distribuição  $\chi^2$  com  $(n-p)$  graus de liberdade.

Como a teoria é elaborada para modelos encaixados, há alguma discordância sobre o uso do teste AIC para modelos não encaixados. Porém, alguns autores o utilizam de uma forma não muito rígida, alegando que seu uso tem importante contribuição para o processo de construção de modelos.

O Critério de Informação de Akaike (AIC) é uma forma de selecionar um modelo a partir de um conjunto de modelos. Para amostras finitas deve-se atentar ao usar este critério que as vezes se torna polarizado e pode falhar na escolha do modelo. O critério é dado pela equação

$$AIC = -2 \sum_{i=1}^n \log L(y_i, \hat{\mu}_i) + 2p$$

O Critério de Informação Bayesiano (BIC), intimamente relacionado com o critério de Akaike (AIC), é utilizado para selecionar modelos pertencentes a um conjunto finito de modelos, baseando-se na função de verossimilhança. O critério pode resolver este problema através de um termo penalizando o número de parâmetros do modelo dado por

$$BIC_{(p)} = 2l(\hat{\beta}_p) + \log(n)$$

e tem como pressuposto a existência de um “modelo verdadeiro” entre os diversos modelos sob seleção. O modelo com o valor de BIC menor é considerado o de melhor ajuste.

Segundo Paula (2004) outro método para avaliar modelos é a análise de deviance que considera as funções de deviance  $D(y; \mu^0)$  e  $D(y; \hat{\mu})$  dos modelos sob  $H_0$  e  $H_1$ , respectivamente, sendo  $\hat{\mu}^0$  a estimativa de máxima verossimilhança sob  $H_0$ . Com isso, a estatística da razão de verossimilhança, considerando  $\phi$  conhecido, é dada por

$$RV = \phi \{D(y; \hat{\mu}^0) - D(y; \hat{\mu})\}$$

sendo que  $RV$  segue assintoticamente uma distribuição  $\chi^2_q$  sendo  $q$  os graus de liberdade do modelo sob a hipótese  $H_0$ . De maneira equivalente pode-se utilizar a estatística  $F$  definida na *procedure* GENMOD como

$$F = \frac{D_1 - D_0}{r \hat{\phi}} \quad (1)$$

sendo  $D_0$  a deviance resultante do ajuste do modelo linear generalizado e  $D_1$  é a deviance do ajuste do submodelo,  $r$  é a diferença do número de parâmetros dos dois modelos e  $\hat{\phi}$  é uma estimativa com base na deviance ou qui-quadrado de Pearson dividido pelo grau de liberdade, cuja distribuição assintótica sobre a hipótese nula é uma  $F_r$ . Na *procedure* GLIMMIX a estatística  $F$  dada por

$$F = \frac{\hat{\beta}' L' (LQL)^{-1} L \hat{\beta}}{(LQL')} \quad (2)$$

sendo  $L$  a matriz para cada efeito que é utilizada para a construção da estatística  $F$ , a matriz  $Q$  depende do método de estimação, por exemplo em modelos lineares generalizados misto  $Q = (X' V(\hat{\theta})^{-1} X)^{-1}$  onde  $V(\theta)$  é a variância marginal da pseudo-resposta.

A estatística da equação 2 possui uma vantagem em relação a estatística de equação 1, pois, a equação 2 não depende do parâmetro de dispersão. A estatística da razão de verossimilhança, quando o parâmetro  $\phi$  é desconhecido, a equação (2)

diferencia-se da equação (1) e segue uma distribuição  $F$  sob a hipótese nula  $H_0$  considerando o número de categorias que estão sendo testadas.

### 3. Resultados

Para avaliar a metodologia proposta neste estudo foram utilizados dados de uma pesquisa com a finalidade de investigar a eficiência de flavonóide no tratamento de “diabetes mellitus”. A população amostral foi constituída de 32 ratos, divididos em quatro grupos denominados Controle, Diabetes, Flavonóides e FLav/Diab e, cada grupo foi subdividido em cinco classes por tamde áreas categorizadas conforme as medidas  $< 200 \mu m^2$ ,  $201$  a  $300 \mu m^2$ ,  $301$  a  $400 \mu m^2$ ,  $401$  a  $500 \mu m^2$  e  $> 500 \mu m^2$ . Coletou-se 100 corpos celulares selecionados conforme as classes de cada rato, extraídos do jejuno de ratos adultos (região do intestino delgado) transferidos para o local de temperatura e iluminação controlada, sendo disponibilizados para o experimento após 90 dias de idade, onde o período experimental obteve 120 dias de duração. Nesta pesquisa a variável resposta de interesse é o número de neurônios que foram distribuídos nas classes.

Todos os procedimentos com animais foram aprovados e realizados em conformidade com os princípios éticos da Sociedade Brasileira de Ciência em Animais de Laboratório (SBCAL) e aprovado pelo Comitê de Ética em Experimentação Animal da Universidade Estadual de Maringá (Parecer: 053/2009).

A Figura 1 apresenta a comportamento do número médio de neurônios e erros-padrão para cada combinação dos níveis de fatores em estudo. Observou-se que, a medida que o valor da média das classes decresceram não ocorreu uma redução proporcional para a variância, implicando assim na sobredispersão. Notou-se que independente dos grupos a maior concentração de neurônios se apresentaram pertinentes as duas classes com menores áreas, e as curvas que representam a distribuição dos neurônios foram aproximadamente iguais.

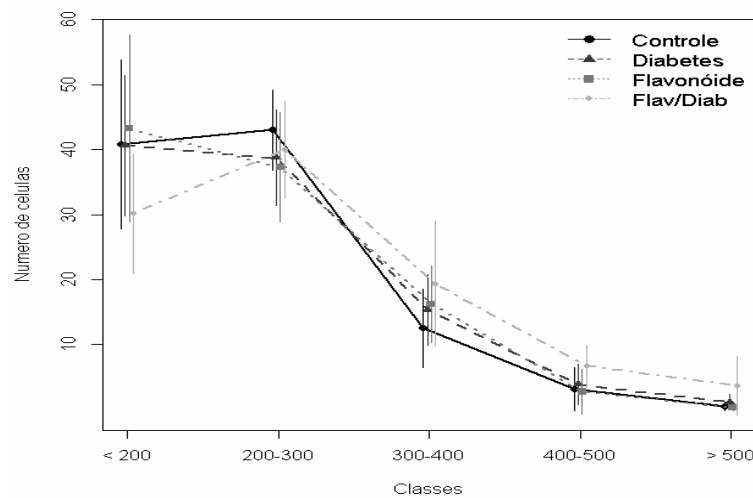


Figura 1 - Médias e erros-padrão do número de neurônios por grupos e classes

Os modelos concorrentes possuem a mesma componente sistemática escrita como

$$\eta_{ijk} = \log(\mu_{ijk}) = \beta_0 + \alpha_i + \gamma_j + (\alpha\gamma)_{ij}$$

sendo  $\beta_0$  o intercepto,  $\alpha_i$  os efeitos dos grupos sendo  $i$  o controle, diabetes, flavonóide e fla/diab e  $\gamma_j$  os efeitos das classes sendo  $j = < 200 \mu m^2, 200 \text{ |----} 300 \mu m^2, 300 \text{ |----} 400 \mu m^2, 400 \text{ |----} 500 \mu m^2, > 500 \mu m^2$  e  $(\alpha\gamma)_{ij}$  o efeito devido a interação, em que  $\kappa$  é o efeito do  $\kappa$ -ésimo rato com  $\kappa = 1, \dots, 32$ . Para obter as estimativas dos efeitos utilizou-se como “baseline” para o fator classe a categoria  $200 \text{ |----} 300 \mu m^2$  e para o fator grupo a categoria Controle.

Na Tabela 1 observa-se os valores das estimativas, que são iguais para os três modelos com seus respectivos erros-padrão. Nota-se que os modelos utilizados para acomodar a sobredispersão apresentaram erros-padrão maiores que o modelo Poisson, isto é, a presença da sobredispersão acarreta em erros-padrão subestimados.

Tabela 1 - Estimativas, Erros-Padrão e P-valores para os três modelos

Trat./Classe	Estimativa	Erros-padrão		
		Poisson	Quase-Poisson	Binomial Negativo
Intercepto	3,7612	0,0539	0,0985	0,1324
Tratamento Flav	-0,1435	0,0791	0,2076	0,1885
Tratamento Diab.	-01041	0,0783	0,1431	0,1882

Tratamento	Flav/Diab		0,0723	0,0841	0,1446	0,2030
Classe	< 200		-0,0537	0,0773	0,2846	0,1877
Classe	301-400		-1,2355	0,1136	0,3784	0,2054
Classe	401-500		-2,6218	0,2071	0,9188	0,2687
Classe	> 500		-4,4543	0,5029	0,1412	0,5312
Trat./Classe	Flav	< 200	0,2031	0,1105	0,0000	0,2660
Trat./Classe	Diab.	< 200	0,1010	0,1108	0,2287	0,2661
Trat./Classe	Flav/Diab	< 200	-0,2284	0,1252	0,2020	0,2898
Trat./Classe	Flav	301-400	0,4059	0,1548	0,5115	0,2872
Trat./Classe	Diab.	301-400	0,3111	0,1558	0,2929	0,2878
Trat./Classe	Flav/Diab	301-400	0,5084	0,1603	0,2828	0,3066
Trat./Classe	Flav	401-500	0,0157	0,3028	1,1072	0,3876
Trat./Classe	Diab.	401-500	0,3192	0,2800	0,4905	0,3700
Trat./Classe	Flav/Diab	401-500	0,8300	0,2685	0,5533	0,3747
Trat./Classe	Flav	> 500	-0,5496	0,8696	0,2024	0,9027
Trat./Classe	Diab.	> 500	0,9150	0,6060	1,0049	0,6525
Trat./Classe	Flav/Diab	> 500	2,0648	0,5500	1,5888	0,6090

A qualidade do ajuste do modelo Poisson foi verificada pelos critérios *scale Deviance* e *scale Pearson* com valores 462.21 e 433.92 respectivamente e os graus de liberdade iguais a 130, apresentados na Tabela 2. A razão dessas *scale* pelos graus de liberdade foram maiores que 1. Com isso, verificou-se que a *scale Deviance* não segue uma distribuição  $\chi^2$ , ou seja, o modelo Poisson não se ajustou bem aos dados de contagem com sobredispersão.

Tabela 2 - Critério de informação de ajuste modelo de Poisson (*proc GENMOD*)

Critérios	Poisson		
	g.l.	Valor	Valor/g.l.
<i>Deviance</i>	130	462,21	3,55
<i>ScaleDeviance</i>	130	462,21	3,55
Pearson $X^2$	130	433,92	3,33
<i>Scale Pearson <math>X^2</math></i>	130	433,92	3,33

Na Tabela 3 através dos critérios *descaleDeviance* e *escale Pearson* observou-se que houve evidências para obter modelos com melhores ajustes ao acomodar a sobredispersão, pois, os valores dos critérios foram menores para os modelos Quase-Poisson e Binomial Negativo em relação ao modelo Poisson, ou seja, os valores estão mais próximo de 1.

Tabela 3 - Critério de informação de ajuste modelos Quase-Poisson e Binomial Negativa

Critérios	Quase-Poisson	Binomial Negativa
-----------	---------------	-------------------

	g.l.	Valor	Valor/g.l.	g.l.	Valor	Valor/g.l.
Deviance	130	462,21	3,55	130	210,20	1,61
Escala Deviance	130	138,47	1,06	130	210,20	1,61
Pearson $X^2$	130	433,92	3,33	130	186,57	1,43
Escala Pearson $X^2$	130	130	1,00	130	186,57	1,43

A Figura 2 apresenta os pontos influentes obtidos através da medida da Distância de Cook quando a variável de interesse segue distribuição de probabilidade de Poisson, sendo que as observações com grandes distâncias de Cook podem exercer um efeito substancial sobre o valor ajustado e, conseqüentemente o modelo de regressão. Neste análise as observações 4, 69 e 150 foram identificadas como pontos influentes.

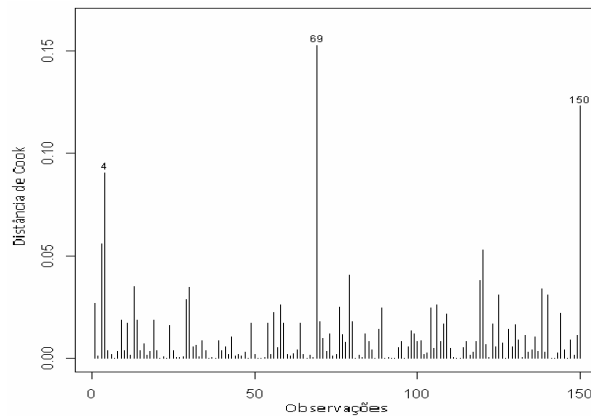


Figura 2 - Distância de Cook

A normalidade dos resíduos foi avaliada pelos gráficos de envelopes apresentado na Figura 3 Para o modelo Poisson verificou-se que os valores residuais excederam as curvas que representam os limites dos valores inferiores e superiores toleráveis, mostrando que os resíduos não seguem a distribuição normal. De outra forma, para os modelos Quase-Poisson e Binomial Negativo os valores residuais seguem aproximadamente uma distribuição normal.



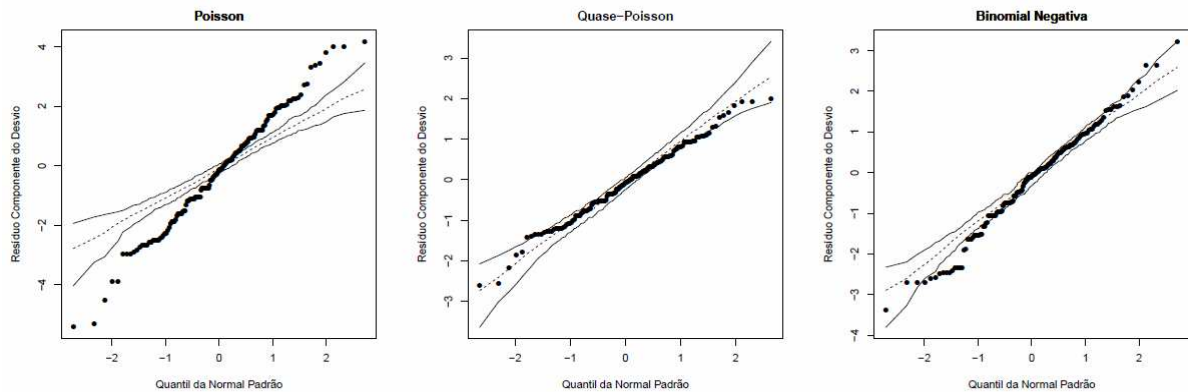


Figura 3 - Envelopes dos três modelos

Pode-se observar os resultados das análises do Tipo 3 na Tabela 4 apresentando valores aproximadamente semelhantes para os parâmetros de escala e dispersão das funções de variância propostas. Para as variáveis predictoras verificou-se que a variável tratamento não foi significativa quando realizou-se os cálculos pela *procedures* GENMOD. Por outro lado, a variável tornou-se significativa quando a análise foi executada pela *procedures* GLIMMIX. Esta divergência dos valores está relacionada com a maneira que as *procedures* abordam as variáveis predictoras.

Tabela 4 - Análise do Tipo III de efeitos fixos.

	GENMOD				GLIMMIX			
	$V(Y) = \phi\mu$		$V(Y) = \mu(1 + \kappa\mu)$		$V(Y) = \phi\mu$		$V(Y) = \mu(1 + \kappa\mu)$	
	Test F	P-Valor	$\chi^2$	P-Valor	Test F	P-Valor	Test F	P-Valor
Classe	58,15	<0,0001	142,80	<0,0001	78,61	<0,0001	131,21	<0,0001
Tratamento	0,36	0,7847	0,62	0,8910	3,52	0,0170	7,10	0,0002
Class/Trat.	1,96	0,0330	37,16	0,0002	1,81	0,0527	2,83	0,0018
	$\hat{\phi} = 1,82279$		$\hat{\kappa} = 0,1171$		$\hat{\phi} = 1,82279$		$\hat{\kappa} = 0,1414$	
	$\phi$		$\kappa$		$\phi$		$\kappa$	

Para avaliar as funções de variância que melhor acomodou a sobredispersão, utilizou-se os critérios de AIC e BIC, porém, não tendo muito rigor quanto aos resultados apresentados na Tabela 5 que apresentou a função de variância da distribuição Binomial Negativa acomodando melhor a sobredispersão.

Tabela 5 - Critérios para avaliação dos modelos

	Poisson	Q-Poisson	Bin. Negativa
AIC	1069,77	1069,77	966,15
BIC	1129,99	1129,99	1029,38

## 4. Discussão

Através dos critérios *scale Deviance* e *scale Pearson* constatou-se a presença de sobredispersão, pois, foram obtidos valores de 3.55 e 3.33 respectivamente, enquanto no caso da distribuição de probabilidade Poisson deveriam aproximar-se de 1, Tabela 2. Por outro lado na Tabela 3 verificou-se valores menores para as *scales* em relação aos valores apresentados pelo modelo Poisson, acarretando eficiência nas estimativas dos parâmetros de escala e dispersão das funções de variância propostas.

Pontos influentes foram identificados utilizando a distância de Cook, que segundo Baltazar (2015) a eliminação de observações que geram preocupações é fundamentada na possibilidade de que, ao retirar uma determinada observação e analisar sua influência nas estimativas, pode ocorrer da observação que esteja influenciando os valores das estimativas esteja sendo induzida por outras observações que não foram identificadas. Estes pontos devem ser avaliados com cautela antes de conhecidos como valores atípicos. Nesta análise os pontos identificados foram 4, 69 e 150, exerceram um peso desproporcional nas estimativas dos parâmetros do modelo. Os pontos foram mantidos, ou seja, para os pesquisadores todas as informações biológicas dos animais são importantes, pois, neste tipo de pesquisa retirar observações pode acarretar em resultados imprecisos.

A Figura 2 é dividida em três gráficos, na primeira observou-se que existem evidências que o modelo de Poisson não se adequou bem aos dados, isto é, a presença da sobredispersão foi um problema para o ajuste deste modelo. Nas outras duas partes percebeu-se a normalidade nos resíduos, ou seja, com a acomodação da sobredispersão utilizando as funções de variância obteve-se modelos com melhores ajustes. Portanto, para esta pesquisa verificou-se que os modelos não foram tendenciosos obtendo as estimativas dos parâmetros de regressão iguais, logo, não faz sentido verificar a distância dos valores observados com os valores preditos.

Na Tabela 4 foram apresentados resultados, obtidos através da *procedure* GENMOD e *procedure* GLIMMIX do *software* SAS 9.3, das análises do Tipo 3 para as

funções de variância em estudo, onde os valores estimados dos parâmetros de dispersão das *procedures* são próximos. Para testar a presença da sobredispersão, verificou-se a hipótese nula  $H_0: \phi = 1$  para a função de variância  $V(Y) = \phi\mu$  e a hipótese nula  $H_0: \kappa = 0$  para função de variância  $V(Y) = \mu(1 + \kappa\mu)$ . Assim, pelo teste da razão de verossimilhança rejeitou-se a hipótese  $H_0: \phi = 1$  obtendo o  $p < 0,0001$  e pelo teste de Wald também rejeitou-se a hipótese  $H_0: \kappa = 0$ , onde o intervalo de confiança para o parâmetro de dispersão foi (0.0730;0.1879), isto é, de fato os dados apresentaram sobredispersão.

Para estimar os parâmetros de regressão estas *procedures* utilizaram os mesmos métodos, porém, não calcularam a estatística F pela mesma expressão. Diante disso, observou-se conforme Tabela 4 uma alteração significativa para o p-valor na variável aleatória tratamento independente da função de variância utilizada. Isto é, a análise realizada pela *procedure* GENMOD o p-valor aponta que o variável aleatório tratamento não é eficiente para tratar de "diabetes mellitus" considerando o nível de significância de 5%, ou seja, não rejeitou-se a hipótese nula. Enquanto, para a *procedure* GLIMMIX a variável tratamento é significativa para o nível de 5% de significância, o que foi importante para o pesquisador. Esta divergência é consequência da forma com que as *procedures* abordam esta variável, isto é, o teste F na *procedure* GENMOD não leva em consideração a categorização da variável, ao passo que, a análise realizada pela *procedure* GLIMMIX considera-se a variável aleatória tratamento como categórica.

Através do diagnóstico foi possível observar que os modelos Quase-Poisson e Binomial Negativo adequaram-se melhor aos dados que o modelo de Poisson. No entanto, não se evidenciou o modelo com melhor ajuste, pois, o teste de Vuong comum para modelos não encaixados não foi utilizado, assim como, os Critérios de AIC e BIC são contestados quando usados para modelos que não são encaixados. Portanto, na sequência deste estudo recomendou-se utilizar Critérios mais eficientes que o de *Deviance*, propondo outras funções de variância e realizar a análise considerando os efeitos aleatórios.

## 5. Apêndice:

Programa em SAS para a análise estatística do Projeto 1.

```
proc genmod data=huc1 plots=all;
class classe (ref='200-300') tratamento(ref='controle') / param=ref;
model freqhuc= classe tratamento classe*tratamento / dist=pois type1 type3
diagnostics;
output out=pois11 pred=predpois1 reschi=res1 resdev=resd ;
```

```
run;
```

```
proc genmod data=huc1 plots=all;
class classe(ref='200-300') tratamento(ref='controle') / param=ref;
model freqhuc= classe tratamento classe*tratamento / dist=Poi pscale type1 type3
diagnostics;
output out=qppois pred=predqppois xbeta=netapoi1 stdxbeta=stdnetapoi1 reschi=res;
run;
```

```
proc genmod data=qppois;
class classe (ref='200-300') tratamento(ref='controle') / param=ref;
model freqhuc= classe tratamento classe*tratamento / dist=nb pscale type1 type3;
output out=qpoisl pred=predpoisl xbeta=netapoi stdxbeta=stdnetapoi reschi=res1;
run;
```

```
proc glimmix data=huc1 ;
class classe tratamento;
variance= _phi*_mu_;
model freqhuc = classe tratamento classe*tratamento/ s link=log;
run;
```

```
proc glimmix data=huc1 ;
class classe tratamento;
variance= _mu_ + _phi*_mu_**2;
model freqhuc = classe tratamento classe*tratamento/ link=log;
run;
```

## 6. Referências

1. ALAM, M. Zero-inflated models for RNA-Seq count data. Tese (Doutorado) — BALL STATE UNIVERSITY, 2015.
2. BESAG, J. Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society. Series B (Methodological)*, JSTOR, p. 192–236, 1974.
3. BONAFEDE, E. et al. Modelling overdispersion heterogeneity in differential expression analysis using mixtures. *arXiv preprint arXiv:1410.8093*, 2014.
4. BOZDOGAN, H. Model selection and akaike's information criterion (aic): The general theory and its analytical extensions. *Psychometrika*, Springer, v. 52, n. 3, p. 345–370, 1987.
5. CHEN, X.; HONG, H.; SHUM, M. Nonparametric likelihood ratio model selection tests between parametric likelihood and moment condition models. *Journal of Econometrics*, Elsevier, v. 141, n. 1, p. 109–140, 2007.
6. DEMÉTRIO, C. G. B. Modelos lineares generalizados em experimentação agrônoma. *Simpósio de estatística aplicada à experimentação agrônoma*, v. 5, p. 1–125, 2002.
7. HINDE, J.; DEMÉTRIO, C. G. Overdispersion: models and estimation. *Computational Statistics & Data Analysis*, Elsevier, v. 27, n. 2, p. 151–170, 1998.
8. HOEF, J. M. V.; BOVENG, P. L. Quasi-poisson vs. negative binomial regression: how

should we model overdispersed count data? *Ecology, Eco SocAmerica*, v. 88, n. 11, p. 2766–2772, 2007.

9. KASSAHUN, W. et al. Marginalized multilevel hurdle and zero-inflated models for overdispersed and correlated count data with excess zeros. *Statistics in medicine, Wiley Online Library*, v. 33, n. 25, p. 4402–4419, 2014.
10. LESTER, R. Overdispersion in marine fish parasites. *Journal of Parasitology, American Society of Parasitologists*, v. 98, n. 4, p. 718–721, 2012.
11. LINDÉN, A.; MÄNTYNIEMI, S. Using the negative binomial distribution to model overdispersion in ecological count data. *Ecology, Eco SocAmerica*, v. 92, n. 7, p. 1414–1421, 2011. Referências 39
12. MCCULLAGH, P.; NELDER, J. A. *Generalized linear models*. [S.l.]: CRC press, 1989. v. 37.
13. MOHEBBI, M.; WOLFE, R.; FORBES, A. Disease mapping and regression with count data in the presence of overdispersion and spatial autocorrelation: a bayesian model averaging approach. *International journal of environmental research and public health, Multidisciplinary Digital Publishing Institute*, v. 11, n. 1, p. 883–902, 2014.
14. MOREL, J. G.; NEERCHAL, N. *Overdispersion models in SAS*. [S.l.]: SAS Institute, 2012.
15. PAULA, G. A. *Modelos de regressão: com apoio computacional*. [S.l.]: IME-USP São Paulo, 2004.
16. QIU, Y.; NORDMAN, D. J.; VARDEMAN, S. B. A pseudo-likelihood analysis for incomplete warranty data with a time usage rate variable and production counts. *IIE Transactions, Taylor & Francis*, v. 46, n. 2, p. 118–130, 2014.
17. RIDOUT, M.; HINDE, J.; DEMÉTRIO, C. G. A score test for testing a zero-inflated poisson regression model against zero-inflated negative binomial alternatives. *Biometrics, Wiley Online Library*, v. 57, n. 1, p. 219–223, 2001.
18. SCHWARZ, G. et al. Estimating the dimension of a model. *The annals of statistics, Institute of Mathematical Statistics*, v. 6, n. 2, p. 461–464, 1978.
19. VARIN, C.; REID, N.; FIRTH, D. An overview of composite likelihood methods. *Statistica Sinica*, v. 21, n. 1, p. 5–42, 2011.
20. VUONG, Q. H. Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica: Journal of the Econometric Society, JSTOR*, p. 307–333, 1989.
21. WOLFINGER, R.; TOBIAS, R.; SALL, J. Computing gaussian likelihoods and their derivatives for general linear mixed models. *SIAM Journal on Scientific Computing, SIAM*, v. 15, n. 6, p. 1294–1310, 1994.
22. WOOLDRIDGE, J. M. Quasi-maximum likelihood estimation and testing for nonlinear

models with endogenous explanatory variables. *Journal of Econometrics*, Elsevier, v. 182, n. 1, p. 226–234, 2014.

23. YU, D.; BAI, P.; DING, C. Adjusted quasi-maximum likelihood estimator for mixed regressive, spatial autoregressive model and its small sample bias. *Computational Statistics & Data Analysis*, Elsevier, v. 87, p. 116–135, 2015.

24. ZHANG, H. et al. A non-parametric model to address overdispersed count response in longitudinal data setting with missingness. *Statistical methods in medical research*, SAGE Publications, p. 0962280215583397, 2015.

Projeto2: Indicadores de referência para a avaliação nutricional específicos para a população Indígena Kaingang-Brasil, 2014.

## Resumo

O estudo objetivou identificar o perfil de saúde e propor indicadores do estado nutricional antropométrico da Comunidade Indígena do Ivaí. A avaliação antropométrica foi realizada estratificando a população em faixas etárias de 0 a <5 anos,  $\geq 5$  a <10 anos,  $\geq 10$  a <20 anos,  $\geq 20$  anos a <60 anos e maiores de 60 segundo índices do SISVAN/OMS. Verificou-se a eficiência dos índices na identificação da obesidade através de uma análise de regressão e em seguida utilizou-se as curvas de sensibilidade e especificidade denominada TG-ROC, com propósito de obter o ponto de corte com melhor desempenho. A população amostral foi constituída de 631 pessoas, sendo 37,40% do sexo masculino e 62,60% do sexo feminino. Os índices score-z e percentil apresentaram-se fidedignos para quase todas as relações de medidas nas faixas etárias. Para a faixa etária de 0 a < 5 anos a relação entre Peso/Altura não se apresentou fidedigna, resultados congruentes nas faixas etárias de  $\geq 5$  a < 10 anos e  $\geq 10$  a < 20 anos sendo as medidas do IMC

confirmadas pelo p-valor=0.0458; 0,0356; 0,0260 com nível de 95% de confiança. Ressalta-se que para as faixas etárias  $\geq 20$  a  $< 60$  anos e  $\geq 60$  anos não existem a diferenciação pelo escore-z e percentil. Os índices de Peso/Idade e Altura/Idade, independente da faixa etária, apresentaram-se adequados e para os outros índices a população passou (observou-se alta prevalência de sobrepeso e obesidade) a apresentar sobrepeso e ser obesa. A análise pelo teste do qui-quadrado de Pearson, evidenciou associação em quase todas as faixas etárias nos dois métodos de classificação. A Razão de Verossimilhança, para os índices P/I (z), P/A(P), IMC(z) e IMC(P) na faixa etária de 0 a  $< 5$  anos de 3,40; 1,08; 1,08 e 1,5 respectivamente, indicou um aumento da chance do indivíduo ser obeso confirmado pela acurácia acima de 0,740. Para as crianças  $\geq 5$  a  $< 10$  anos os índices de P/I(z) ; P/I (P), a Razão de Verossimilhança negativa sugeriu que o baixo peso está aumentado com ocorrência de falsos positivos. Apesar de obter alguns índices com eficiência no desempenho para a classificação da obesidade confirmada pelos testes de diagnósticos e de associação é necessário ampliar a amostra populacional associados com testes mais robustos.

Palavras-chave: Avaliação Nutricional, Obesidade, Antropometria, Índios Kaingang.

## Abstract:

The study aimed to identify the health profile and propose indicators of the anthropometric nutritional status of the Indigenous Community of Ivaí. The anthropometric assessment was performed stratifying the population into age groups second indexes SISVAN / WHO. It has been found the efficiency of identification indexes of obesity through a regression analysis and then we used the sensitivity and TG-ROC curves specificity for obtaining the cutoff point with better performance. The population consisted of 631 people, with 37.40% males and 62.60% females. The z-score and percentile indexes were presented reliable for almost all relations measures in the age groups. For age from 0 to <5 years the relationship between weight / height did not appear reliable, consistent results for  $\geq 5$  to <10 and  $\geq 10$  to <20 being the BMI measures confirmed by the p-value= 0.0458; 0.0356; 0.0260. For those aged  $\geq 20$  to <60 and  $\geq 60$  no differentiation by z-score and percentile. Rates of Weight / Age and height / age, regardless of age, had to be appropriate and others indexes the population was classified as overweight and obese. The analysis by Pearson's chi-square, showed association in almost all age groups in both classification methods. The Likelihood Ratio for the indices W / A (z), P / A (P), BMI (z) and BMI (P) in the age group 0 to <5 years of 3.40; 1.08; 1.08 and 1.5



respectively, indicated an increased chance of being obese individual confirmed by the accuracy above 0.740. For children  $\geq 5$  to  $<10$  years, rates of W / A (z); P / I (P), negative likelihood ratio suggests that low birth weight is increased with false positives. While getting some indexes with efficient performance for obesity classification confirmed by diagnostic and association tests is necessary to expand the sample population associated with more robust testing.

**Key words:** Nutritional Assessment, obesity, anthropometry, Kaingang Indians.

## 1. Introdução

Atualmente o perfil de saúde dos povos indígenas é influenciado pela introdução de novas doenças e pela delimitação arbitrária de territórios, dificultando desta forma a utilização do solo para subsistência. Acrescentam-se também as perseguições e mortes que ocorreram indiscriminadamente, atingindo muitas vezes populações inteiras (SANTOS e COIMBRA JR., 2003). Outrossim, a saúde indígena é caracterizada pela elevada prevalência de doenças infecto-parasitárias e desnutrição, principalmente na infância, além de enfrentar desafios da nova agenda de saúde pública, tais como a obesidade, doenças crônicas metabólicas, alcoolismo e violência. As sociedades indígenas experimentam ainda problemas relacionados ao esgotamento de recursos naturais, à contaminação ambiental e à dificuldade de acesso a gêneros alimentícios diversificados.

Em relação à saúde de povos indígenas situados em outras partes da América, em especial no Canadá e nos EUA, os estudos de GUGELMIN e SANTOS (2001) e (WHO, 2008) enfatizaram o aumento da prevalência da obesidade, da hipertensão arterial e do "diabetes mellitus" tipo II. Tais estudos consideraram que a obesidade e as doenças crônicas não transmissíveis apresentaram-se como um dos principais problemas de morbimortalidade nas sociedades, em contraposição às

doenças infecto-parasitárias, que desde a segunda metade do século XX revelam uma menor relevância no cenário epidemiológico. A situação observada nas populações indígenas norte-americanas não se aplica à grande maioria das sociedades indígenas no Brasil, nas quais as doenças infecto-parasitárias encontram-se francamente endêmicas. No entanto, desde a década de 70, estudos com etnias indígenas localizadas em diferentes estados brasileiros têm relatado incremento na prevalência de obesidade e doenças associadas (CARDOSO et al., 2001; GUGELMIN e SANTOS, 2001).

Além do contexto epidemiológico, a heterogeneidade presente entre as etnias indígenas exerce papel primordial na avaliação do processo saúde-doença destas populações. No Brasil, existem aproximadamente 220 povos indígenas, falantes de 180 línguas, revelando elevada diversidade ecológica, sócio-cultural, política e econômica, reflexo de constituições sociais e trajetórias históricas singulares (RICARDO, 2002).

Há alguns grupos que vivem ainda em relativo isolamento na Amazônia, no entanto, existem etnias com grande parte de seu contingente populacional urbanizado. Esta diversificação inviabiliza análises genéricas em relação ao perfil de saúde e nutrição de sociedades indígenas, considerando principalmente as várias estratégias destes povos em relação à utilização dos recursos naturais e à capacidade de produção e aquisição de gêneros alimentícios. Todavia, através da compilação de estudos com grupos específicos, é possível identificar claramente a precariedade das condições de saúde e nutrição, além da desigualdade e posição de desvantagem das populações indígenas em relação a outros segmentos da sociedade nacional (COIMBRA JR. e SANTOS, 2000).

Os povos indígenas estão expostos a transformações ambientais e sócio-econômicas, que os colocam em situação de alta vulnerabilidade frente aos problemas de ordem alimentar e nutricional (PÍCOLI et al., 2006). Disponibilidade monetária, estado nutricional e situação de acesso aos alimentos são frequentemente utilizados como indicadores de insegurança alimentar desta população (MONTEIRO, 2003).

Os determinantes dos agravos nutricionais nos indígenas estão relacionados a inúmeras transformações de ordem demográfica, socioeconômica, cultural e ambiental. Em muitas comunidades indígenas a subsistência a partir da caça, coleta e pesca estão sendo modificadas em decorrência de novos regimes econômicos e

da diminuição dos limites territoriais levando a deficiência dietética (GUGELMIN e SANTOS, 2001).

As informações disponíveis sobre o processo da transição nutricional para os povos indígenas no Brasil são limitados, pois este segmento populacional não é incluído em inquéritos nacionais de saúde e nutrição. Observe-se, no entanto, que diversos estudos epidemiológicos voltados para a avaliação nutricional de populações indígenas concordam com a elevada prevalência da desnutrição crônica em crianças e do aumento da obesidade entre os adultos (LOURENÇO et al. 2008, WELCH et al. 2009, ABRASCO, 2009).

O Inquérito Nutricional de Saúde e Nutrição dos Povos Indígenas realizado em 2008/2009 por meio de amostras representativas de crianças e mulheres em idade reprodutiva das grandes macrorregiões do país apresentaram resultados importantes sobre o excesso de peso e níveis pressóricos elevados nas mulheres e elevada proporção de crianças com déficit de crescimento e moderada prevalência de baixo peso para a idade e a quase inexistência de déficits nutricionais pelo índice peso para estatura (ABRASCO, 2009).

Informações sobre o estado nutricional são de extrema importância para a avaliação das condições de saúde de coletividades humanas, principalmente ao se considerar o processo de transição nutricional e epidemiológica que vêm ocorrendo mundialmente (WHO, 1995). Em países desenvolvidos, o aumento na prevalência de sobrepeso na população já vem sendo acompanhado há algumas décadas, questionando-se acerca das implicações deste incremento sobre a morbimortalidade futura, principalmente no que tange às doenças crônicas não transmissíveis.

Diante do exposto este estudo objetivou a identificar o perfil de saúde e nutrição da população indígena e propor indicadores de referência para a avaliação nutricional específico para a população Indígena Kaingang-do Ivaí.

## 2. Metodologia

Trata-se de um estudo analítico de corte transversal, realizado na Comunidade Indígena Kaingang do Ivaí em Manoel Ribas- PR. Os Kaingang encontram-se entre as maiores etnias indígenas no país, distribuídos em 28 terras

indígenas, somando entre 22 e 25 mil pessoas nos Estados de São Paulo, Paraná, Santa Catarina e Rio Grande do Sul (MOTA et al., 2003).

A terra indígena Ivaí é uma das 19 áreas indígenas existentes no Paraná com uma área aproximada de 7.800 hectares, localizada a 6 km da cidade de Manoel Ribas (distância entre a aldeia e o centro da cidade). Em 2002 era constituída por 287 famílias, num total de 1.085 pessoas. Neste mesmo ano, 57.9% das pessoas possuíam até 19 anos de idade, denotando uma população jovem (MOTA et. al., 2003). Em junho do ano de 2010 a população estimada nesta Comunidade era de 1420 pessoas, segundo informações da FUNAI/FUNASA.

As idades das crianças, adolescentes e adultos foram coletadas de um livro de registro de nascimento disponível na UBS localizada dentro da aldeia indígena. As medidas de peso e estatura (ou comprimento) foram obtidas segundo as recomendações da Organização Mundial da Saúde (OMS-2006, 2008). Os indivíduos foram pesados portando vestimentas leves. Crianças menores de dois anos foram medidas no posto de saúde da aldeia, com o auxílio de um agente de saúde indígena, devidamente treinado. Os demais indivíduos foram examinados no posto ou em seus próprios domicílios e neste último caso sempre acompanhados pelo agente indígena. Crianças de até dois anos foram pesadas em uma balança pediátrica mecânica com capacidade máxima de 16kg e precisão de 0,01kg, e seu comprimento aferido em decúbito dorsal, com um antropômetro. Indivíduos mais velhos foram pesados em uma balança eletrônica com capacidade máxima de 150kg e precisão de 0,1kg, e a estatura medida em posição ortostática, com os pés juntos e descalços, utilizando-se o antropômetro com a mesma precisão. As gestantes não foram avaliadas.

A classificação do estado nutricional foi realizada de acordo com os critérios adotados pelo SISVAN (Sistema de Vigilância Alimentar e Nutricional) do Ministério da Saúde através das Orientações para a Coleta e Análise de Dados Antropométricos em Serviços de Saúde (2011) que recomenda a classificação para as crianças menores de 5 anos pela Norma de referência da WHO (2006) e para as crianças a partir de 5 anos pela Norma de Referência da WHO (2008). É importante ressaltar que as curvas de avaliação do crescimento para crianças dos 5 aos 19 anos foram reanalisadas do NCHS (National Center Statistics Health) de 1977, além do alisamento das curvas no período de transição entre os menores de 5 anos de

idade, avaliados segundo o estudo-base dos dados lançados em 2006 e dos indivíduos a partir de 5 anos.

Para o estudo foi considerado o erro de estimativa de 5%, confiabilidade e precisão da amostra de 95%, adotando-se o percentual de 50%, que fornece o tamanho amostral máximo, obtendo-se o número de 302 pessoas. Para o sorteio foram acrescentados 20% ao total, levando-se em consideração eventuais perdas que poderiam ocorrer. Dessa forma, o tamanho amostral final estabelecido deveria ser de 362 pessoas, no entanto, observou-se a participação voluntária de indígenas de várias faixas etárias e pela cultura indígena não poderíamos recusar a sua inclusão na amostra totalizando desta forma 631 indivíduos.

A partir do número amostral estipulado, foi realizada uma estratificação proporcional por faixa etária, pois este procedimento é necessário para a obtenção dos indicadores nutricionais e de obesidade. As faixas etárias foram divididas de  $0 < 5$  anos,  $\geq 5$  a  $< 10$  anos,  $\geq 10$  a  $< 20$  anos,  $\geq 20$  a  $< 60$  anos e  $\geq 60$  anos para a comparabilidade com outros estudos, conforme a proposta do SISVAN (BRASIL, 2008 e OMS-2006).

Os índices utilizados para as crianças de  $0$  a  $< 5$  anos foram o Peso para Idade (P/I), Peso para Altura (P/A), Índice de Massa Corporal para a Idade (IMC/I) e Altura para Idade (A/I), classificados segundo os critérios estabelecidos no Manual de Orientações do SISVAN(2004). Para as crianças com idade de  $\geq 5$  a  $< 10$  anos o índice de peso para altura (P/A) não é realizado e para os adolescentes de  $\geq 10$  a  $< 20$  anos, os índices de P/I e P/A também não são realizados. Para a população adulta ( $\geq 20$  anos a  $< 60$  anos), o diagnóstico nutricional foi realizado através do IMC ( $\text{Kg}/\text{m}^2$ ) utilizando os pontos de corte propostos pela WHO(1995), sendo  $\text{IMC} < 18,5$   $\text{Kg}/\text{m}^2$ , considerado baixo peso;  $\text{IMC} \geq 18,5$   $\text{Kg}/\text{m}^2$  a  $< 25,0$   $\text{Kg}/\text{m}^2$  normal; de  $25,0$   $\text{Kg}/\text{m}^2$  a  $< 30,0$   $\text{Kg}/\text{m}^2$ , sobrepeso e  $\geq 30,0$   $\text{Kg}/\text{m}^2$  obeso. Para os idosos ( $\geq 60$  anos) foi utilizado um critério diferenciado devido as alterações fisiológicas nos idosos como o declínio da altura, diminuição do peso com a idade, alterações ósseas, mudanças na quantidade do tecido adiposo e redução da massa muscular, com pontos de corte  $\leq 22$   $\text{Kg}/\text{m}^2$ ,  $\geq 22,0$   $\text{Kg}/\text{m}^2$  a  $< 27,0$   $\text{Kg}/\text{m}^2$  e  $\geq 27,0$   $\text{Kg}/\text{m}^2$ .

Assim, para avaliar a fidedignidade dos índices score-z e percentilempregou-se o método de calibração baseado na análise de regressão. Nesta análise uma modificação da curva ROC, denominada TG-ROC, demonstrou as variações de sensibilidade e especificidade segundo os índices antropométricos em estudo. A

rotina computacional para gerar curvas TG-ROC foi programada no software SAS, e cabe ressaltar que foi utilizado como padrão ouro o SISVAN.

Para construir uma nova classificação mais consistente para a comunidade indígena foi utilizado o método K-Means que é utilizado para identificar *clusters* homogêneos objetivando novos pontos de corte para a população indígena. Este método consiste em cada iteração avaliar a variância dentro e entre os *clusters*. A distância entre dois clusters é calculada como a distância média entre todos os pares de medidas nos dois diferentes *clusters*, ou seja, a distância entre dois clusters é definida como a diferença entre dois centróides. Esta abordagem é amplamente utilizada para procedimento de agrupamentos, no qual salienta sua utilização em estudos como de Black et al. (2013) que examinaram as características antropométricas identificando indivíduos pelo método K-Means. Khanna et al. (2011) compararam registros de dietas por uma distância métrica usando o *kernel* K-Means *clustering* para identificar padrões alimentares temporais, Frankenfeld et al.(2015) que investigaram a prevalência de diabetes e obesidade em pessoas que frequentam lugares com opções de alimentos mais saudáveis e lugares com opções de alimentos menos saudáveis identificadas pelo método K-Means. O algoritmo é finalizado quando todos os indivíduos da população estão classificados

A pesquisa foi aprovada pelo Comitê Permanente de Ética em Pesquisa Envolvendo Seres Humanos da UEM (COPEP)/Comissão Nacional de Ética em Pesquisa (CONEP) Parecer 412.427/2014), e Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq - Projeto 474283/2011-9), obtendo-se autorização para entrada para coleta dos dados em Terra indígena pelo Ministério da Justiça/ Fundação Nacional do Índio (Protocolo 09/AAEP/PRES/2014). O Termo de Consentimento Livre e Esclarecido (TCLE) foi apresentado para todos os participantes e após a anuência e assinaturas foram coletados os dados.

### 3. Resultados

Dos 631 indígenas Kaingang, 236(37,40%) eram do sexo masculino e 395(62,60%) do sexo feminino. A distribuição para os extratos de idade conforme as recomendações do SISVAN estão apresentadas na Tabela 1.

Tabela 1 - A distribuição da população segundo a faixa etária, sexo e idades máxima,

mínima e média±DP. Maringá, 2014.

Faixa Etária	Total N (%)	Feminino N (%)	Masculino N (%)	Idade		
				Mínima	Máxima	Média ± DP
0 a < 5	100 (15,8)	50 (12,7)	50 (21,2)	22 dias	4,99	2,88 ± 1,36
≥ 5 a < 10	121 (19,2)	79 (20,0)	42 (17,8)	5,4 anos	9,99	8,52 ± 1,41
≥ 10 a < 20	168 (26,6)	112 (28,3)	56 (23,7)	10,09 anos	19,98	14,01 ± 2,91
≥ 20 a < 60	232 (36,8)	150 (39,0)	82 (34,8)	20 anos	59,61	33,18 ± 10,18
≥ 60	10 (1,6)	4 (1,0)	6 (2,5)	60 anos	79,6	68,80 ± 6,44
<b>Total</b>	<b>631</b>	<b>395 (62,6)</b>	<b>236 (37,4)</b>			

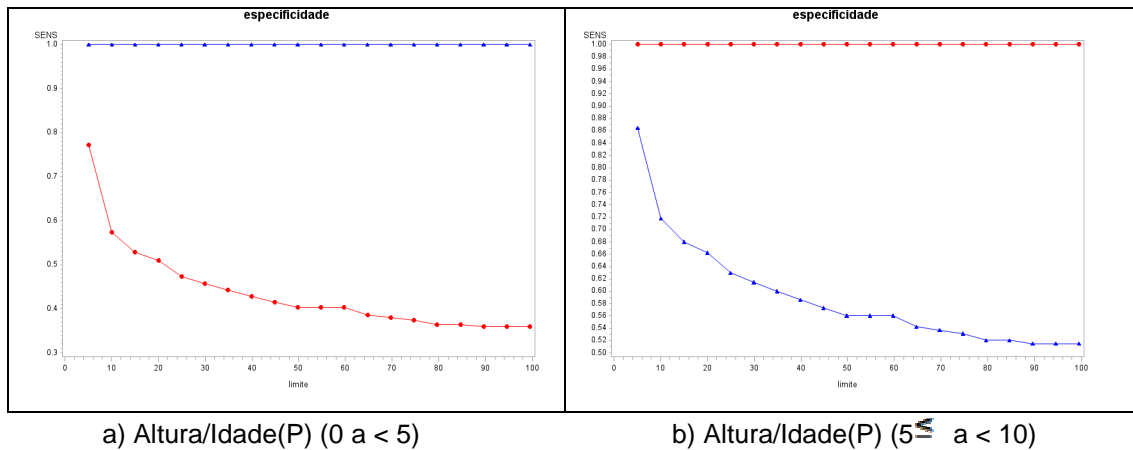
A análise foi realizada em duas etapas, a primeira para verificar se os índices eram eficientes para a determinação da obesidade e na etapa seguinte identificar pela sensibilidade e especificidade o ponto de corte com melhor desempenho. Assim, para avaliar a fidedignidade dos índices score-z e percentil empregou-se o método de calibração baseado na análise de regressão, testando se o coeficiente angular da reta era igual a 1. Uma vez identificada a fidedignidade dos índices para cada nível da faixa etária detectou-se o ponto de corte para obesidade nesta população.

De acordo com a Tabela 2, os índices score-z e percentil foram calibrados através de uma análise de regressão apresentando-se fidedignos, ou seja, sendo convergente nos resultados, consistentes e com precisão para quase todas as relações de medidas nas faixas etárias. Para a faixa etária de 0 a < 5 anos apenas a relação entre P/A (Peso/Altura) apresentou evidência de não ser fidedigna, resultado observado também para as faixas de ≥ 5 a < 10 anos e ≥ 10 a < 20 anos nas relações de medidas do IMC. Diante deste resultado os índices score-z e percentil foram analisados separadamente.

Foram obtidas curvas de sensibilidade e especificidade para identificar adequadamente os obesos e não obesos. Verificou-se ainda que o ponto encontrado através da curva TG-ROC não identifica a doença em estudo na população, pois, em alguns índices os indivíduos foram classificados como obesos e em outros como não obesos, destacando-se que nestas classificações existem grandes quantidades de falsos positivos e falsos negativos. Para alguns índices não foi possível obter os pontos de corte através da curva TG-ROC, como por exemplo, para o índice percentil das medidas Altura/Idade nas faixas etárias de 0 a < 5 e ≥ 5 a < 10 apresentadas nas Figuras 1a e 1b respectivamente. Na Figura 1a a curva de sensibilidade se mantém constante no valor de 100% apontando que todos os indivíduos são obesos, o que não ocorre na Figura 1b, onde a curva de

especificidade se mantém constante em 100% identificando que todas as pessoas são não obesas.

Figura 1 - Curvas de sensibilidade e especificidade(TG-ROC) para pontos de corte.



Para verificar qual índice apresentava melhor desempenho observou-se para a idade de 0 a < 5 anos que as medidas de Peso/Idade, Altura/Idade e IMC podem ser utilizadas tanto no escore-z quanto no percentil diferentemente das medidas de Peso/Altura para menores de 5 anos, e IMC para idades  $\geq 5$  a < 10 anos e  $\geq 10$  a < 20 anos confirmados na Tabela 2 pelo p-valor ( $p=0.0458$ ;  $0,0356$ ;  $0,0260$ ) com nível de 95% de confiança. Ressalta-se que para as faixas etárias de  $\geq 20$  a < 60 anos e  $\geq 60$  anos não apresenta a diferenciação pelo escore-z e percentil.

Tabela 2 - Calibração dos índices antropométricos (percentil e escore-z). Maringá-PR, 2014.

Idade	Índice de Calibração	Estimativa ( $\beta$ )	Erro Padrão	p-valor
0 a < 5	P/I(z) (97) vs P/I(P) (84)	1.26	0.030	0.0738
	P/A(z) (97) vs P/A(P) (84)	2.08	0.078	0.0458
	A/I(z) (98) vs A/I(P) (76)	1.19	0.049	0.1597
	IMC(z) (97) vs IMC(P) (84)	1.87	0.088	0.0637
$\geq 5$ a < 10	P/I(z) (94) vs P/I(P) (93)	1.06	0.020	0.2063
	A/I(z) (94) vs A/I(P) (85)	1.92	0.083	0.0571
	IMC(z) (94) vs IMC(P) (90)	2.09	0.061	0.0356
$\geq 10$ a < 20	A/I(z) (159) vs A/I(P) (148)	1.59	0.065	0.0709
	IMC(z) (159) vs IMC(P)(157)	2.40	0.056	0.0260

(z)=escore z; (P)=percentil



Pela Tabela 3 observou-se que a maioria da população segundo os índices de Peso/Idade e Altura/Idade, independente da faixa etária, apresentaram-se adequados, no entanto, na análise do Peso/Altura e IMC observou-se que a população passava a ser sobrepeso e obesa para todas as faixas etárias analisadas. Este fato nos faz lembrar que a grande maioria das pessoas da comunidade Kaingang apresentam uma altura abaixo dos padrões normais, quando comparada a população brasileira. Para a população  $\geq 20$  a  $< 60$  anos e  $\geq 60$  anos esta análise se apresentou mais complexa, pois, para estas faixas etárias não existe outro índice relacionado à idade e sexo, obtendo-se somente o IMC.

Os índices observados na Tabela 3 foram obtidos a partir do padrão SISVAN, destacando que a estatura desta população indígena é um aspecto influente na avaliação da calibragem dos índices score-z e percentil, isto é, existe uma discrepância na interpretação dos resultados fornecidos por estes índices quando os pesos dos indivíduos estão relacionados com a altura.

Tabela 3 - Distribuição da população nas faixa etária, segundo a classificação dos índices. Maringá-PR. 2014.

Idade	Índice(total)	Baixo peso		Adequado		Sobre peso		Obeso	
		N	%	N	%	N	%	N	%
0 a < 5	Peso/Idade(z)(97)	3	3.1	86	88.7	---	---	8	8.2
	Peso/Idade(P)(84)	36	42.9	48	57.1	---	---	---	---
	Peso/Altura(z)(97)	---	---	2	2.1	82	84.5	13	13.4
	Peso/Altura(P)(84)	---	---	---	---	74	74.0	10	10.0
	Altura/Idade(z) (98)	44	44.9	54	55.1	---	---	---	---
	Altura/Idade(P) (76)	27	35.5	49	64.5	---	---	---	---
	IMC (z)(97)	---	---	2	2.1	82	84.5	13	13.4
	IMC (P)(84)	---	---	2	2.4	82	97.6	---	---
$\geq 5$ a < 10	Peso/Idade (z)(117)	5	4.3	112	95.7	---	---	---	---
	Peso/Idade (P)(116)	4	3.4	112	96.6	---	---	---	---
	Altur/Idade(z)(117)	40	34.2	77	65.8	---	---	---	---
	Altur/Idade(P)(105)	39	37.1	66	62.9	---	---	---	---
	IMC (z) (117)	---	---	83	71.0	27	23.0	7	6.0
	IMC (P) (111)	---	---	83	74.8	26	23.4	2	1.8
$\geq 10$ a < 20	Altur/Idade(z)(159)	63	39.6	96	60.4	---	---	---	---
	Altur/Idade(P)(148)	67	45.3	81	54.7	---	---	---	---
	IMC (z)(159)	---	---	116	73.0	32	20.1	11	6.9
	IMC (P) (157)	---	---	116	73.9	28	17.8	13	8.3
$\geq 20$ a < 60	IMC (196)	6	2.5	56	24.1	73	31.4	61	26.3

$\geq 60$	IMC (10)	2	2.0	4	4.0	2	2.0	---	---
-----------	----------	---	-----	---	-----	---	-----	-----	-----

(z)=escore z; (P)=percentil

Diante do exposto optou-se por construir uma nova classificação mais consistente para a comunidade indígena e, desta forma foi utilizado o método K-Means para identificar pontos de cortes com melhor eficácia (Tabela 4).

Assim para verificar se a classificação do SISVAN e a classificação obtida pelo K-means estariam descrevendo a população indígena, foi realizada a análise pelo teste do qui-quadrado de Pearson, observando-se que em quase todas as faixas etárias os testes evidenciaram associação entre os dois métodos de classificação. Ressalta-se que os índices P/I e A/I para as faixas etárias de 0 a < 5 anos,  $\geq 5$  a < 10 anos e  $\geq 10$  a < 20 anos os indígenas foram classificados para a baixa estatura ou baixo peso (Tabela 4).

Observou-se ainda que na classificação pelo K-means os indivíduos se dividem de forma proporcional, diferentemente da classificação do SISVAN observando que poucos indivíduos são classificados para os extremos. Destaca-se ainda que o método K-Means identifica grupos homogêneos de tal forma que a variabilidade dentro da categoria seja mínima e entre categorias seja máxima. Este método categoriza os dados em números K de *clusters*, onde cada aglomerado destes possui uma nova média denominada de centróide e cada observação é selecionada pela distância euclidiana entre o centróide e o valor observado, isto é, a observação pertence ao cluster que possui à média mais próxima de sua medida.

Tabela 4 - Classificação da população segundo os índices antropométricos, valores críticos do SISVAN, K-means e p-valor. Maringá-PR, 2014.

IDADE	ÍNDICES	CLASSIFICAÇÃO	VALORES CRÍTICOS SISVAN	VALORES CRÍTICOS K-MEANS	p-valor
0 a < 5	Peso/Idade (z) (97)	Baixo Peso	[ ≥ -3 a < -2) (3)	[ ≥ -2.44 a < -0.49) (38)	< 0.0000
		Peso Adequado	[ ≥ -2 a < 2) (86)	[ ≥ -2.44 a < -0.49) (45)	
		Peso Elevado	[ > 2) (8)	[ ≥ 1.47) (14)	
	Peso/Idade (P) (84)	Baixo Peso	[ ≥ 0.1 a < -3) (36)	[ ≥ 0.70 a < 45.40) (36)	< 0.0001
		Peso Adequado	[ ≥ 3 a ≤ 97) (48)	[ ≥ 45.40] (48)	
	Peso/Altura (z) (97)	Eutrofia	[ ≥ -2 a < -1) (2)	[ ≥ -1.17 a < 0.55) (27)	< 0.0000
Risco de Sobrepeso		[ ≥ -1 a < 2) (64)	[ ≥ 0.55 a < 1.70) (34)		
Sobrepeso Obesidade		[ > 2 a ≤ 3) (18) [ > 3) (13)	[ ≥ 1.70 a < 4.03) (29) [ ≥ 4.03) (7)		
Peso/Altura (P) (84)	Risco de Sobrepeso	[ ≥ 15 a ≤ 97) (54)	[ ≥ 31.9 a < 61.7) (12)	< 0.0000	
	Sobrepeso	[ > 97 a ≤ 99.9) (20)	[ ≥ 61.7 a < 86.3) (24)		
	Obesidade	[ > 99.9) (10)	[ ≥ 86.30) (48)		
Altura/Idade (z) (98)	Muito baixa estatura	[ < -3) (22)	[ ≥ -4.89 a < -2.64) (31)	< 0.0000	
	Baixa Estatura	[ ≥ -3 a < -2) (22)	[ ≥ -2.64 a < -0.94) (44)		
	Estatura Adequada	[ ≥ -2) (54)	[ ≥ -0.94) (23)		
Altura/Idade (P) (76)	Baixa Estatura	[ ≥ 0.10 a < -3) (27)	[ ≥ 0.10 a < 23.4) (61)	< 0.0013	
	Estatura Adequada	[ ≥ -3) (49)	[ ≥ 23.40) (15)		
≥ 5 a < 10	Peso/Idade (z) (117)	Muito Baixo Peso	[ < -3) (1)	[ ≥ -3.74 a < -0.94) (34)	0.0125
		Baixo Peso	[ ≥ -3 a < -2) (4)	[ ≥ -0.94 a < 0.25) (63)	
		Peso Adequado	[ ≥ -2 a < 2) (112)	[ ≥ 0.25) (20)	
	Peso/Idade (P) (116)	Baixo Peso	[ ≥ 0.10 a < -3) (4)	[ ≥ 1.2 a < 43.4) (34)	0.1899
Peso Adequado		[ ≥ 3 a ≤ 97) (112)	[ ≥ 43.4 a < 96.9) (82)		
Altura/Idade (z) (117)	Muito Baixa Estatura	[ < -3) (12)	[ > -9.59) (4)	0.0000	
	Baixa Estatura	[ ≥ -3 a < -2) (28)	[ ≥ -9.59 a < -1.63) (60)		
	Adequada	[ ≥ -2 a < 3) (77)	[ ≥ -1.63 a < 0.60) (53)		
Altura/Idade (P) (105)	Baixa Estatura	[ ≥ 0.10 a < 3) (39)	[ ≥ 0.3 a < 20.8) (19)	0.0002	
	Adequada	[ ≥ 3 a < 99.9) (66)	[ ≥ 20.8 a < 72.6) (86)		
≥ 10 a < 20	Altura/Idade (z) (159)	Muito Baixa Estatura	[ < -3) (11)	[ ≥ -14.5 a < -2.15 ) (59)	0.0000
		Baixa Estatura	[ ≥ -3 a < -2) (52)	[ ≥ -2.15 a < 0.32) (99)	
		Estatura Adequada	[ ≥ 2 a > 3) (96)	[ ≥ 0.32 ) (1)	
Altura/Idade (P) (148)	Baixa Estatura	[ ≥ 0.1 a < 3) (67)	[ ≥ 0.2 a < 19.8) (22)	0.0000	
	Estatura Adequada	[ ≥ 3) (81)	[ ≥ 19.8) (126)		

Na análise dos dados da Tabela 5 as crianças de 0 a < 5 anos apresentaram índices P/I (z), P/A(P), IMC(z) e IMC(P) com a Razão de Verossimilhança maior que 1 indicando que o resultado positivo aumenta a chance do indivíduo ser positivo, com resultados de 3,40; 1,08; 1,08 e 1,5 respectivamente. A proporção de acertos, ou seja, a acurácia demonstra que o total de indivíduos seja verdadeiramente

positivos ou negativos e alguns índices apresentaram resultados acima de 0,740 sugerindo que o índice se apresenta como um bom indicador para o diagnóstico da obesidade ou desnutrição na comunidade indígena. Todas as análises apresentaram significância estatística.

Na análise dos dados das crianças  $\geq 5$  a  $< 10$  anos destacam-se os índices de P/I(z) ; P/I (P), com razão de verossimilhança negativos sugerindo que o baixo peso está aumentado, e nestes casos pode existir muitos falsos positivos. Para esta faixa etária a Especificidade do IMC afasta a ocorrência da obesidade em 100% da população. Em todas as análises houve significância estatística dos testes.

Na análise dos dados dos adolescentes  $\geq 10$  anos e  $< 20$  anos destacam-se o IMC para a classificação da obesidade e da A/I para o baixo peso para o escore z e percentil. O IMC utilizado para a classificação de adultos  $\geq 20$  anos a  $< 60$  anos e para os  $\geq 60$  anos se apresentaram adequados para a classificação da obesidade com sensibilidade de 100.

Tabela 5 - Distribuição da população com valores dos testes diagnósticos. Maringá-PR, 2014.

Idade	Índice	N	SENS	ESP	P	VP+	VP-	RV+	RV-	p-val	ACUR
0 a < 5	P/I (z)	97	57.1	100.0	8.24	100.0	93.4	----	0.428	< 0.0001	0.938
	P/I(P)	84	60.0	82.3	35.7	83.3	58.3	3.40	0.487	< 0.0001	0.690
	P/A(z)	97	100.0	7.4	72.1	73.7	100.0	1.08	0.000	< 0.0001	0.742
	P/A(P)	84	41.6	100.0	35.7	100.0	22.8	----	0.580	< 0.0001	0.500
	IMC(z)	97	100.0	8.0	74.2	75.7	100.0	1.08	0.000	< 0.0001	0.762
	IMC(P)	84	100.0	33.3	92.8	95.1	100.0	1.50	0.000	< 0.0001	0.952
	A/I(z)	98	41.3	100.0	31.6	100.0	34.3	0.41	0.580	< 0.0001	0.551
	A/P(P)	76	44.2	100.0	35.5	100.0	30.6	----	0.557	< 0.0013	0.552
$\geq 5$ a $< 10$	P/I(z)	117	51.5	100.0	4.27	100.0	17.1	----	0.940	0.0125	0.213
	P/I(P)	116	4.87	100.0	3.44	100.0	30.3	----	0.950	0.1899	0.672
	P/A(z)	----	----	----	----	----	----	----	----	----	----
	P/A(P)	----	----	----	----	----	----	----	----	----	----
	IMC(z)	117	56.6	100.0	29.0	100.0	68.7	----	0.430	< 0.0001	0.777
	IMC(P)	111	37.8	100.0	25.2	100.0	49.5	----	0.620	< 0.0001	0.585
	A/I(z)	117	62.5	100.0	34.1	100.0	68.8	----	0.370	< 0.0001	0.794
	A/P(P)	105	45.3	100.0	34.1	100.0	28.8	----	0.540	< 0.0002	0.444
$\geq 10$ a 20	A/I(z)	159	39.8	100.0	37.5	100.0	1.04	----	0.600	< 0.0001	0.380
	A/I(P)	148	53.1	100.0	45.3	100.0	27.1	----	0.460	< 0.0001	0.601
	IMC (z)	159	100.0	100.0	37.0	100.0	100.0	----	0.000	< 0.0001	1.000
	IMC(P)	157	32.8	100.0	35.3	100.0	27.6	----	0.670	< 0.0001	0.464
$\geq 20$ a 60	IMC	196	100.0	64.5	51.0	74.6	100.0	2.82	0	< 0.0001	0.826

≥60 anos	IMC	9	100.0	85.7	22.2	66.6	100.0	7.0	0	<0.0001	0.888
----------	-----	---	-------	------	------	------	-------	-----	---	---------	-------

SENS= sensibilidade;ESP= especificidade; P= prevalência; VP+= valor preditivo positivo; VP-= valor preditivo negativo; RV+= razão de verossimilhança positiva; RV-= razão de verossimilhança negativa; p-val= nível de significância; ACUR= acurácia.

#### 4. Discussão

Os Kaingans constituem uma das duas maiores populações indígenas do sul do Brasil, pertencem ao grupo linguístico Gê, são seminômades, vivem nos planaltos e praticam atividades básicas como a agricultura, a caça e a pesca. Apesar de passarem por processos de aculturação, possuem costumes ainda muito diferentes das demais pessoas da população brasileira não indígena (UTIYAMA, et al., 2000).

O objetivo desta pesquisa de analisar o perfil de saúde e nutrição da comunidade indígena Kaingang do Ívai, foi pontuado nos estudos realizados por Lourenço et al. (2008); Welch et al. (2009); ABRASCO (2009) que observaram um crescente aumento na prevalência da desnutrição crônica em crianças e uma percentagem elevada de indivíduos adultos obesos, destacando que a obesidade está correlacionada com o fator de risco cardiovascular (COSTA-FONT J, GIL J, 2005 e CERCATO et al.,2004), e que esta ocorrência esta presente principalmente nas mulheres, devido o contato com o meio urbano (RIBAS et al.,2007).

Os estudos do estado nutricional realizados com as populações indígenas de crianças Suruí, Xavante e Wari (ORELLANA et al., 2006; LEITE et al., 2006; 2007), apontam para elevadas prevalências de desnutrição. Esses resultados, quando comparados aos da Pesquisa Nacional de Demografia e Saúde da Criança e da Mulher (BRASIL, 2008), que utilizou as curvas da OMS (WHO, 2006), revelaram diferenças importantes, destacando a baixa estatura para a idade, seis vezes maior nas crianças indígenas em relação às não indígenas. Essas constatações são condizentes com a presente pesquisa e com os respectivos perfis de saúde das populações indígenas estudadas, marcados por elevadas taxas de mortalidade infantil e de morbidade por doenças infecciosas e parasitárias (ORELLANA et al., 2006; LEITE et al., 2006; 2007; ORELLANA et al., 2007). Diversos pesquisadores afirmam que o sobrepeso e a obesidade se apresentam como um problema

decorrente de países subdesenvolvidos e de populações de baixa renda como as populações indígenas (SILVA et al., 2005; SAWAIA et al.,1995; PELEGRINI et al.,2010; STEYN et. al., 2005).

Nesta pesquisa, a amostra populacional constituída por 37,4% de indivíduos do sexo masculino e 62,6% do sexo feminino com o maior percentual na faixa etária de  $\geq 20$  a  $< 60$  anos (36,8%), adotou-se os índices de medidas antropométricas *score-z* e percentil recomendados pelo SISVAN para a análise dos dados. Durante o processo de análise, verificou-se a existência de incompatibilidade entre estes índices na descrição dos dados, assim, para as medidas antropométricas relacionadas ao peso e altura (P/A) na faixa etária de 0 a  $< 5$ , observou-se que estes índices não se apresentavam consistentes e precisos, ou seja, não descreviam a população de mesma forma, assim como a medida do IMC nas faixas etárias de  $\geq 5$  a  $< 10$  e  $\geq 10$  a  $< 20$ . Com isto, os índices foram usados separadamente na identificação dos obesos e não obesos através da construção das curvas TG-ROC, observando que em torno de 45% dos índices utilizados, os resultados identificavam todos como obesos ou todos como não obesos, o que não é normal em uma população, destacando-se diferenças importantes entre as frequências dos descritores nutricionais.

Os estudos de Seal e Kerac, (2007); e de De Onis et al. (2006), apresentaram que as variações nas prevalências seguiram a tendência de aumento da baixa estatura para a idade e de redução do baixo peso para a idade ao se empregar a WHO/2006, o que tem sido também registrado em outros estudos com amostras não-indígenas. Esta incompatibilidade também foi observada no presente estudo entre os índices *score-z* e percentil para a identificação dos indivíduos obesos sugerindo que este resultados poderiam estar relacionado com a estatura desta população. Desta forma, optou-se por uma reclassificação que apresentasse melhor desempenho ao selecionar a população, em que recorreu-se ao método K-Means que identificou os obesos e não obesos na comunidade indígena com novos pontos de corte.

Na Tabela 2 observa-se a fidedignidade entre os pontos de corte recomendados pelo SISVAN e os pontos de corte obtidos pelo método K-Means, avaliados através do teste de hipótese do teste qui-quadrado de Pearson, onde a hipótese nula testa a igualdade entre os índices *score-z* e percentil com nível de confiança de 95%, demonstrando que a relação entre as medidas peso e idade (P/I)

na faixa etária de  $5 \geq a < 10$  anos são fidedignas na descrição da comunidade indígena Kaingang. Nas demais relações de medidas antropométricas, independente da faixa etária, os valores críticos do SISVAN e valores críticos obtidos pelo uso do método K-Means não são fidedignos, ou seja, rejeita-se a hipótese nula com nível de confiança de 95%. Acrescente-se que uma medida fidedigna é consistente e precisa, pois, fornece a medida estável da variável.

Para faixa etária de 0 a  $< 5$  anos os pontos de corte do método K-Means identificaram tanto as pessoas obesas quanto as não obesas na relação entre o peso e a idade (P/I) para o índice *escore-z* e, outra relação que apresenta identificação para obesos e não obesos é o peso com a altura (P/A) tanto para o *escore-z* quanto para percentil.

Para as medidas de IMC verificou-se que os pontos de corte do método K-means e os pontos de corte do SISVAN, analisados pela estatística do teste qui-quadrado de Pearson, não são fidedignos para todas as faixas etárias com exceção da faixa etária  $\geq 60$ , onde o número de indivíduos nesta faixa etária é baixo. No entanto, para as medidas de IMC os pontos de corte obtidos pelo método K-Means obtiveram melhor desempenho na identificação dos indígenas que os pontos do SISVAN.

Nas crianças indígenas, os casos de sobrepeso foram encontrados a partir do índice P/A e IMC para idade, mostrando-se mais sensível quando a altura estava presente, resultando em importantes alterações nas prevalências de sobrepeso. Mesmo diante da não disponibilidade de estudos com dados empíricos provenientes do conjunto de curvas de crescimento WHO/2005, estabelecendo modelos de risco que associem sobrepeso infantil com obesidade subsequente ou outras doenças e agravos, chama atenção o aumento das estimativas de sobrepeso geradas pelo IMC em relação àquelas do índice P/A para ambas as curvas.

Os dados de sobrepeso apontam para duas situações bastante distintas, que encontram continuidade nos perfis dos adultos e adolescentes destas mesmas etnias, conforme destacado nos estudos de Leite et al. (2006); Orellana et al. (2007; Lourenço et al.(2008) que observaram entre os Wari' que o sobrepeso e a obesidade constituem ocorrências raras, enquanto entre os Suruí e os Xavante as prevalências chegam a ser alarmantes.

Boretto et al. (2015) avaliaram o estado nutricional das comunidades indígenas Guarani e Kaingang do Estado do Paraná, que apresentaram elevadas

prevalências de sobrepeso e obesidade, sendo a maior freqüência entre as mulheres, o mesmo referenciado por Castro et al. (2012) que também destacaram alta prevalência de obesidade abdominal entre os adolescentes indígenas Kaingangs.

O estudo transversal realizado por Dos Anjos, et al. (2011) com medidas antropométricas para avaliar a prevalência de síndrome metabólica, diabetes mellitus e obesidade entre as populações Kaingang nativos da região central do Estado do Paraná, obteve resultados negativos quanto a prevalência de doenças nessas comunidades.

Se para as crianças brasileiras há diversos estudos nacionais que permitem caracterizar a trajetória da desnutrição, como é o caso do Estudo Nacional da Despesa Familiar (ENDEF), da Pesquisa Nacional sobre Saúde e Nutrição (PNSN), da Pesquisa Nacional de Demografia e Saúde (PNDS), Pesquisa de Orçamentos Familiares (POF), entre outras (CONDE e GIGANTE, 2008), para as crianças indígenas o que se conhece da situação nutricional deriva de estudos de casos em etnias específicas (SANTOS e COIMBRA, 2003; LEITE et al., 2007).

Para os vários inquéritos nacionais, que apresentam bases de dados consolidados, é factível proceder reanálises a partir do novo conjunto de curvas de crescimento, o que permite a geração de resultados comparáveis. No caso das crianças indígenas, as bases estão dispersas, em geral de posse dos pesquisadores que realizaram os estudos. Com isto, é possível que, com a adoção da WHO/2005, se perca a comparabilidade com os estudos anteriores sobre a situação nutricional das crianças indígenas, que em sua ampla maioria se apoiaram na NCHS/1977. Este é um sério problema, já que o monitoramento da situação nutricional das crianças indígenas é de fundamental importância, considerando, em particular, a vulnerabilidade que essas populações, em acelerados processos de mudanças socioeconômicas e ambientais, apresentam no que tange à dimensão nutricional (LEITE et al., 2007).

Essencialmente, os resultados deste estudo revelam e reforçam a marcante desigualdade frente aos indicadores nutricionais registrados entre os não indígenas, coexistindo nesse grupo desordens nutricionais características do processo de transição nutricional tardia, em que são ainda importantes as prevalências de retardo estatural na infância e adolescência e o excesso de peso crescente a partir da infância. Espera-se que o esclarecimento do quadro nutricional



dos indígenas Kaingáng no contexto escolar auxilie o norteamento e a reflexão de políticas públicas adequadas que se voltem à promoção alimentar e nutricional direcionada ao segmento.

A interpretação dos dados antropométricos a luz da realidade social, sanitária e ambiental a que estão submetidas grande parte das populações indígenas que vivem no Brasil, parece não deixar dúvidas de que o perfil de crescimento, bem como os agravos nutricionais em menores de 10 anos, é estritamente relacionado às condições insalubres de saneamento e moradia, assistência à saúde insuficiente e situações de insegurança alimentar, esta determinada pela falta de recursos monetários para aquisição de alimentos e pela falta de área física para a produção (GARNELO et al., 2003; FÁVARO et al., 2007; LEITE et al., 2007).

Entende-se que a próxima etapa investigativa dos agravos nutricionais na população Kaingang, seja o afunilamento das informações, incluindo assim aspectos referentes aos padrões alimentares e de atividade física.

## 5. Conclusão

A estatura desta população indígena é um aspecto influente na avaliação da calibragem dos índices score-z e percentil, isto é, existe uma discrepância na interpretação dos resultados fornecidos por estes índices quando os pesos dos indivíduos estão relacionados com a altura. Desta forma, os índices propostos pelo SISVAN para avaliação de medidas antropométricas apresentaram-se insuficientes para esta análise diante dos resultados obtidos, ou seja, para a população indígena Kaingang Ívai não foi possível definir como pessoas obesas e não obesas tomando como base os índices do SISVAN.

O ideal seria realizar um estudo populacional para atestar os índices preconizados pela OMS/SISVAN, uma vez que é necessário conhecer a mediana de cada idade e sexo segundo o indicador, no entanto, como não foi possível o estudo populacional, a pesquisa foi realizada através de uma amostra com poucos indígenas em cada idade e sexo, assim, optou-se a identificar os indivíduos pelos pontos de corte obtidos através do método K-Means, considerando como padrão-ouro os índices que OMS/SISVAN propõem. Ressalva que não se pode concluir para uma tomada de decisão baseada nestes novos pontos de corte.

Apesar de obter alguns índices com eficiência no desempenho para a classificação da obesidade confirmada pelos testes de diagnósticos e de associação é necessário ampliar a amostra populacional associados com testes mais robustos.

## 6. Apêndice:

Programa em SAS para a análise estatística do Projeto 2.

```
## Calibração dos índices escore-z e percentil.

proc means data=taque.de0a4comindices;
var BA__ BAZ;
run;
data taque.de0a4comindices;
set taque.de0a4comindices;
bazpad=(BAZ-1.8759)/1.86;
imcpad=(IMC__-83.2511)/18.68;
padraouroimc=(IMC__<15)*0+(IMC__>=15)*1;
padraourobaz=(BAZ<-1)*0+(BAZ>=-1)*1;
if (IMC__='.') then padraouroimc='.';
if (BAZ='.') then padraourobaz='.';
run;

proc reg data=taque.de0a4comindices;
model imcpad=bazpad;
run;
proc iml;
p=(1-probt(9.9612,1))*2;
print p;
quit;

## Construção das curvas de sensibilidade e especificidade para identificar os
pontos de corte.

proc sort data=taque.de5a9comindices1;
by BA__;
run;

data roc;
set taque.de5a9comindices1;
drop i;
array imcper[20]imcper1-imcper20;
do i=1 to 20 ;
imcper[i]=(BA__<28.10+3.59*i)*0+
(BA__>=28.10+3.59*i)*1;
end;
run;

%macro mytabela(m);
%do i=1 %to &m;
title "tabela imcper&i";
proc sort data=roc;
by descending imcper&i descending padraouroimc;
run;
proc freq data= roc order=data;
tables padraouroimc*imcper&i;
```

```

title 'ordenado';
run;

proc freq data= roc;
tables padraouroimc*imcper&i;
title 'nao ordenado';
run;

proc freq data=roc order=data;
tables padraouroimc ;
exact binomial;
where imcper&i=1;
output out=sens&i binomial;
title 'sensibilidade';
run;

proc freq data=roc ;
tables padraouroimc;
exact binomial;
where imcper&i=0;
output out=esp&i binomial;

title 'especificidade';
run;

%end;

%mend;

%mytabela(20 );

data senstotal;
keep SENS;
set Sens1-sens20;
SENS=_BIN_;
run;
data esptotal;
keep ESP;
set espl-esp20;
ESP=_BIN_;
run;
DATA sensesp;
merge senstotal esptotal;
run;
data taque.sensesp;
set sensesp;
id=_n_;
um_esp=1-esp;

limite=28.10+3.59*id;

run;symbol1 interpol=join
value=dot;
ods graphics on;

proc gplot data=taque.sensesp;
plot sens*um_esp;
run;
ods graphics off;
symbol1 color=red
interpol=join
value=dot
height=1;
symbol2 font=marker value=C
color=blue
interpol=join
height=.5;

```

```
proc gplot data=taque.sensesp;
  plot sens*limite esp*limite / overlay ;
run;
```

## 6. Referências

1. Santos RV, Coimbra CEA. Cenários e tendências da saúde e da epidemiologia dos povos indígenas no Brasil. In: Coimbra CEA, Santos RV, Escobar AL, organizadoras. Epidemiologia e Saúde dos Povos Indígenas no Brasil. Rio de Janeiro: Editora Fiocruz, Abrasco; 2003.
2. Gugelmin AS, Santos RV. Ecologia humana e antropometria nutricional de adultos Xavante, Mato Grosso, Brasil. CadSaúdePública 2001; 17(2):313-322.
3. World Health Organization. Worldwide prevalence of anaemia 1993-2005: WHO global database on anaemia. Geneva: World Health Organization; 2008.
4. Cardoso AM, Mattos IE, Koifman RJ. Prevalência de fatores de risco para doenças cardiovasculares na população Guaraní-Mbyá do Estado do Rio de Janeiro. Cad Saúde Pública 2001; 17:345-354.
5. Ricardo CA. Povos Indígenas no Brasil, 1996-2000. São Paulo: Instituto Socioambiental; 2002.
6. Coimbra CEA, Santos RV. Saúde, minorias e desigualdade: algumas teias de inter-relações, com ênfase nos povos indígenas no Brasil. Ciência e Saúde Coletiva 2000, 5(1): 125-132.
7. Picoli RP, Carandina L, Ribas DLB. Saúde materno-infantil e nutrição de crianças Kaiowá e Guaraní, Área Indígena de Caarapó, Mato Grosso do Sul, Brasil. Cad Saúde Pública 2006; 22(1): 223-227.
8. Monteiro CA. Fome, Desnutrição e Pobreza: além da Semântica. Saúde e Sociedade. 2003; 12(1): 7-11.
9. Lourenço AE, Santos RV, Orellana JD, Coimbra Jr CE. Nutrition transition in Amazonia: obesity and socioeconomic change in the Suruí Indians from Brazil. Am J Hum Biol. 2008; 20:564-71.
10. Welch JR, Ferreira AA, Santos RV, Gugelmin AS, Werneck G, Coimbra Jr. CEA. Nutrition Transition, Socioeconomic Differentiation and Gender Among Adult Xavante Indians, Brazilian Amazon. Hum. Ecology. 2009; 37:13–26.
11. Associação Brasileira de Saúde Coletiva. I Inquérito Nacional de Saúde e Nutrição dos Povos Indígenas. Rio de Janeiro: 2009. Relatório final, nº 7. Funasa.

12. World Health Organization. Physical Status: The Use and Interpretation of Anthropometry. Technical Report Series 854. Geneva: World Health Organization; 1995.
13. Mota LT. Diagnostico etno-ambiental da Terra Indígenalvai - PR. Programa Interdisciplinar de Estudos de Populações. Maringá: 2003. UEM/LAAE.
14. Brasil. Ministério da Saúde. Secretaria de Atenção à Saúde. Departamento de Atenção Básica-Coordenação Geral da Política de Alimentação e Nutrição-Vigilância Alimentar e Nutricional-SISVAN. Orientações para Coleta e Análise de Dados Antropométricos e Serviços de Saúde, 2008.
15. World Health Organization. WHO child growth standards: length/height-for-age, weight-for-age, weight-for-length, weight for height and body mass index-for-age: methods and development. [serial on line] [cited 2015 ago 01]. Available from: [http://www.who.int/childgrowth/standards/Technical\\_report.pdf](http://www.who.int/childgrowth/standards/Technical_report.pdf).
16. Black MH, Watanabe RM, Trigo E, Takayanagi M, Lawrence JM, Buchanan TA, Xiang AH. High-Fat Diet Is Associated with Obesity-Mediated Insulin Resistance and  $\beta$ -Cell Dysfunction in Mexican Americans<sup>1,2,3</sup>. *J Nutr*. 2013; 143(4): 479–485.
17. Khanna N, Eicher-Miller HA, Boushey CJ, Gelfand SB, Delp EJ. Temporal Dietary Patterns Using Kernel k-Means Clustering. *ISM*. Author manuscript; available in PMC 2014. *ISM*. 2011: 375–380.
18. Frankenfeld CL, Leslie TF, Makara MA. Diabetes, obesity, and recommended fruit and vegetable consumption in relation to food environment sub-types: a cross-sectional analysis of Behavioral Risk Factor Surveillance System, United States Census, and food establishment data. *BMC Public Health*. 2015; 15: 491.
19. Utiyama SRDR, Guardiano J, Petzl-Erler ML, Mocelim V, Messian-Reason JT. Perfil de auto-anticorpos em índios das tribos Kaingang e Guarani do Sul do Brasil. *Rev Panam Salud Publica* 2000; 7(6):371-376.
20. Costa-Font J, Gil J. Obesity and the incidence of chronic diseases in Spain: a seemingly unrelated probit approach. *Econ Hum Biol*. 2005; 3(2):188-214.
21. Cercato C, Mancini MC, Arguelli AMC, Passos Q, Villares SMF, Halpern A. Systemic hypertension, diabetes mellitus, and dyslipidemia in relation to body mass index: evaluation of a Brazilian population. *Rev Hosp Clin*. 2004; 59 (3): 113-8.
22. Orellana JD, Coimbra Jr CE, Lourenço AE, Santos RV. Nutritional status and anemia in Suruí Indian children, Brazilian Amazon. *JPediatr (Rio J)*. 2006;82:383-8.
23. Leite MS, Santos RV, Coimbra Jr CE, Gugelmin AS. Alimentação e nutrição dos povos indígenas no Brasil. In: Kac G, Sichieri R, Gigante DP, organizadores. *Epidemiologia Nutricional*. Rio de Janeiro: Fiocruz/Atheneu; 2007. p. 503-17.

24. Orellana JD, Basta PC, Santos RV, Coimbra Jr CE. Morbidade hospitalar em crianças indígenas Suruí menores de dez anos, Rondônia, Brasil: 2000 a 2004. *Rev Bras Saude Mater Infant.*2006;7:281-7.
25. Seal A, Kerac M. Operational implications of using 2006 World Health Organization growth standards in nutrition programmes: secondary data analysis. *BMJ.*2007;334:733.
26. Sawaya AL, Dallal G, Solymos G, De Sousa MH, Ventura ML, Roberts Sb et al. Obesity and malnutrition in a Shantytown population in the city of São Paulo, Brazil. *Obes Res* 1995;3 (Supl 2):107s-15.
27. Pelegrini A, Silva DA, Petroski EL, Gaya AC. Overweight and obesity in seven to nine-year-old Brazilian students: data from the Brazilian Sports Project.*Rev Paul Pediatr* 2010;28:290-5.
28. Steyn NP, Labadarios D, maunder E, Nel J, Lombard C. Directors of the National Food Consumption Survey. Secondary anthropometric data analysis of the National Food Consumption Survey in South Africa: the double burden. *Nutrition*2005;21:4-13.
29. Conde WL, Gigante DP. Epidemiologia da desnutrição infantil. In: Kac G, Sichieri R, Gigante DP, organizadores. *Epidemiologia nutricional*. Rio de Janeiro, RJ: Fiocruz/Atheneu; 2007. p.281-95.
30. De Onis M, Onyango AW, Borghi E, Garza C, Yang H; WHO MulticentreGrowthReferenceStudyGroup. Comparison of the World Health Organization (WHO) Child Growth Standards and the National Center for Health Statistics/WHO international growth reference: implications for child health programmes. *Public Health Nutr.* 2006;9:942-7.
31. World Health Organization. WHO Child Growth Standards: length/height-for-age, weight-for-age, weight-for-length, weight-for-height and body mass index-for-age. Methods and development. WHO (nonserial publication). Geneva: WHO; 2006.
32. Boaretto JD, Molena-Fernandes CA, Pimentel GGA. Estado nutricional de indígenas Kaingang e Guarani no estado do Paraná, Brasil. *Ciênc. saúdecoletiva.* 2015; 20(8):1413-8123.
33. De Castro, TG, Barufaldi LA, Schlussek MM, Conde WL, Leite, MS, Schuch I. Waist circumference and waist circumference to height ratios of Kaingang indigenous adolescents from the State of Rio Grande do Sul, Brazil. *Cad Saúde Pública.* 2012; 28(11): 2053-2062.
34. Dos Anjos HNK, Toledo MJD, Mota LT, Previdelli ITS, Dos Anjos AF, Saruhashi TR, Carrara MA, Batista MR. Prevalence of Metabolic Syndrome among Kaingang Native Americans in Southern Brazil. *Brazilian Arch of boil and tech.* 2011; 54(1): 81-89.

35. Conde WL, Gigante DP. Epidemiologia da desnutrição infantil. In: Kac G, Sichieri R, Gigante DP, organizadores. Epidemiologia nutricional. Rio de Janeiro, RJ: Fiocruz/Atheneu; 2007. p.281-95.
36. Garnelo L; Sampaio. Bases sócio-culturais do controle social em saúde indígena. Problemas e questões na Região Norte do Brasil. Cad Saúde Pública 2003; 19 (1): 311-318.
37. Fávaro T, Ribas DLB, Zorzatto JR, Segall-Corrêa AM, Panigassi G. Segurança alimentar em famílias indígenas Teréna, Mato Grosso do Sul, Brasil. Cad Saúde 2007; 23(4): 785-793.
38. Ribas, Dulce L. B.; Fávaro, Thatiana; Martins, Camila da S.; Dias, Lazaro A.; Samúdio, Helio; Stolte, Vania P. Revistada Sociedade Brasileira de Alimentação e Nutrição Sobre peso e Obesidade em Mulheres Indígenas.vol. 32 / n. Suplemento 9º Congresso Nacional da SbanOut. 2007.

## TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO

### Título do Projeto: **Perfil de Saúde e Nutrição em Comunidade Indígena Kaingang do Ivaí**

Os pesquisadores abaixo relacionados vêm através deste convidá-lo a participar do projeto de pesquisa acima referido que será desenvolvido com a população da Terra Indígena do Ivaí, no Estado do Paraná. Você será esclarecido pelo pesquisador e/ou pelo agente indígena de saúde, em relação à sua participação. **O projeto tem como objetivo identificar o perfil de saúde e nutrição da Comunidade Indígena do Ivaí e os fatores associados aos agravos nutricionais.**

Você responderá a um questionário com informações sobre aspectos socioeconômicos, nutricionais e estilo de vida.

Será submetido a uma avaliação do estado nutricional com medidas de peso, altura e circunferência da cintura e do quadril para identificar seu estado nutricional.

Será coletado sangue da ponta do seu dedo com uma micro lanceta, que servirá para investigar se você está com anemia.

Poderá receber orientação sobre alimentação saudável e alimentos ricos em ferro, para prevenir a anemia.

**A PARTICIPAÇÃO DA PESQUISA É VOLUNTÁRIA.** Você tem o direito de não concordar em participar ou mesmo de retirar-se do estudo em qualquer momento, sem riscos para o seu tratamento médico. Se você desejar e concordar em participar, deve assinar na linha abaixo deste documento em duas vias, sendo uma para o pesquisador responsável, e a outra ficará em seu poder.

#### Riscos e Benefícios

**Os riscos para a saúde são muito pequenos e a coleta de sangue poderá causar apenas algum desconforto, que durará poucos segundos. A investigação será sempre realizada com a presença de um intermediário nomeado pela comunidade.** Caso você ache inapropriada alguma das questões que constam do questionário ou lhe produza sentimentos indesejáveis, poderá interromper a entrevista a qualquer momento. Os benefícios para as pessoas envolvidas na pesquisa serão: o encaminhamento dos casos de obesidade mórbida, de desnutrição ou de anemia para avaliação clínica e tratamento pela equipe de saúde da Unidade Básica de Saúde local; sugestões de medidas profiláticas para controle dos agravos nesta população.

Eu, \_\_\_\_\_ (no caso de menor de idade, o responsável) \_\_\_\_\_ após ter lido e entendido as informações e esclarecido todas as dúvidas referentes a este estudo com os professores Dra Taqueco Teruya Uchimura e Dr Lúcio Tadeu Mota, concordo voluntariamente e dou meu consentimento, sem ter sido submetido a qualquer tipo de pressão ou coação

\_\_\_\_\_ Data: \_\_\_\_/\_\_\_\_/\_\_\_\_\_  
Assinatura (do pesquisado ou responsável) ou impressão datiloscópica.

Eu, Prof. Dra Taqueco Teruya Uchimura declaro que forneci todas as informações a respeito do estudo ao paciente.

\_\_\_\_\_ Data: \_\_\_\_/\_\_\_\_/\_\_\_\_\_  
Assinatura

#### Equipe (Incluindo pesquisador responsável):

- 1- Nome: Dra Taqueco Teruya Uchimura \*\* (44)3261-4513,  
Endereço: Av. Colombo, 5790 - Bloco I-90, Maringá, PR.
- 2- Nome: Lucio Tadeu Mota Telefone: \*\* (44)3261- 4670.  
Endereço: Av. Colombo, 5790 - Bloco I-90, Maringá, PR.
- 3- Nome: Emerson Barili



4. Nome: Adriana Mayumi Moriwaki

Qualquer dúvida ou maiores esclarecimentos procurar um dos membros da equipe do projeto ou o Comitê Permanente de Ética em Pesquisa Envolvendo Seres Humanos (COPEP) da Universidade Estadual de Maringá – Sala 01 – Bloco 010 – Campus Central – Telefone: \*(44)3261-4444.