REVISTA DE ARQUEOLOGIA

Volume 32 No. 1 2019

O USO DO PROGRAMA R NA ARQUEOLOGIA PAULISTA: UM ESTUDO DE CASO DOS GRUPOS CERAMISTAS TUPIGUARANI E ITARARÉ-TAQUARA

Glauco Constantino Perez*, Mercedes Okumura**, Marisa Coutinho Afonso***, Lucio Tadeu Mota****

RESUMO

Este artigo apresenta um estudo de caso explorando o amplo potencial de aplicação do programa de estatística R na análise de dados arqueológicos. O programa R é um ambiente de desenvolvimento integrado para cálculos estatísticos e gráficos, tendo seu uso a partir de pacotes que são bibliotecas de cálculos aplicadas a áreas de estudo específicas. Neste artigo, utilizam-se dados dos sítios arqueológicos dos grupos ceramistas das tradições Tupiguarani e Itararé-Taquara, que se localizam nas bacias dos Rios Tietê e Paranapanema no Estado de São Paulo, como estudo de caso para a utilização do método. Apresentam-se análises descritivas de atributos elencados previamente e análises inferenciais com modelos de regressão logística de Distância Euclidiana (DE) e Regressão logística Bayesiana.

Palavras-chave: Programa R; Tradição Tupiguarani; Tradição Itararé-Taquara.

- * Graduado em História pela Universidade Estadual de Maringá (2007), Mestre em História pelo Programa de Pós-graduação em História Linha de Pesquisa: Fronteira e Bens Culturais da Universidade Estadual de Maringá (2010). Doutor pelo Programa de Pós-graduação em Arqueologia da Universidade de São Paulo, Museu de Arqueologia e Etnologia MAE/USP (2018). Pós-doutorando pelo Instituto de Biociências (IB/USP). E-mail: glauco1113@gmail.com. https://orcid.org/0000-0002-1472-3360
- ** Professora Doutora do Instituto de Biociências da Universidade de São Paulo e coordenadora do Laboratório de Estudos Evolutivos Humanos (LEEH/IB/USP). E-mail: okumura@ib.up.br. https://orcid.org/0000-0002-1894-6430
- *** Professora Associada do Museu de Arqueologia e Etnologia da Universidade de São Paulo MAE/USP e coordenadora do Laboratório de Arqueologia da Paisagem e Geoarqueologia MAE/USP. E-mail: marisa.afonso@usp.br. https://orcid.org/0000-0002-4397-8856
- **** Professor Doutor do Departamento de História (DHI/UEM) e coordenador do Laboratório de Arqueologia, Etnologia e Etnohistória da Universidade Estadual de Maringá (LAEE/UEM/Maringá). Email: https://orcid.org/0000-0001-7064-1389.

DOI: https://doi.org/10.24885/sab.v32i1.617

THE USE OF R PROGRAM IN THE SÃO PAULO'S ARCHAEOLOGY: A CASE STUDY WITH TUPIGUARANI AND ITARARÉ-TAQUARA CERAMIST GROUPS

ABSTRACT

This paper presents a case study exploring the high potential of applying the statistical program R to analyze archaeological data. The R program is a programming language and software environment for statistical computing and graphics and its use is based on packages or libraries applied to specific areas of study. In this paper we use data from archaeological sites of ceramic groups of the Tupiguarani and Itararé-Taquara traditions that were located at the Tietê and Paranapanema Rivers basins at São Paulo State as a case study for illustrating the method. We present descriptive analysis of previously selected attributes and inferential analysis with logistic regression models of Euclidian Distance (ED) and Bayesian Logistic Regression.

Keywords: R software; Tupiguarani Tradition; Itararé-Taquara Tradition.

EL USO DEL PROGRAMA R EN LA ARQUEOLOGÍA PAULISTA: UN ESTUDIO DE CASO DE LOS GRUPOS CERAMISTAS TUPIGUARANI E ITARARÉ-TAQUARA

RESUMEN

Este artículo presenta un estudio de caso explorando el amplio potencial de aplicación del programa de estadística R en el análisis de datos arqueológicos. El programa R es un entorno de desarrollo integrado para cálculos estadísticos y gráficos, con su uso a partir de paquetes que son bibliotecas de cálculos aplicados a áreas de estudio específicas. En este artículo se utilizan datos de los sitios arqueológicos de los grupos ceramistas de las tradiciones Tupiguarani e Itararé-Taquara que se ubican en las cuencas de los Ríos Tietê y Paranapanema en el Estado de São Paulo como estudio de caso para la utilización del método. Se presentan análisis descriptivos de atributos previamente enumerados y análisis inferenciales con modelos de regresión logística de Distancia Euclidiana (DE) y Regresión logística Bayesiana.

Palabras clave: Programa R; Tradición Tupiguarani; Tradición Itararé-Taquara.

INTRODUÇÃO

Neste artigo apresentam-se os estudos estatísticos realizados sobre os dados coletados a respeito dos grupos das tradições arqueológicas Tupiguarani e Itararé-Taquara para os sítios arqueológicos localizados entre as bacias dos Rios Tietê e Paranapanema, no estado de São Paulo, Brasil. De forma mais específica, a intenção deste artigo é utilizar a análise dos dados levantados durante a pesquisa de Perez (2018) para demonstrar o potencial do Programa R na arqueologia, bem como, através dele, apresentar possíveis áreas de contato e fronteiras culturais entre esses grupos ceramistas.

Nosso conhecimento a respeito da chegada dos grupos Tupiguarani e Itararé-Taquara na região sul e sudeste do Brasil no Holoceno tardio é baseado principalmente nos últimos 40 anos de pesquisas em sítios arqueológicos nessas regiões. Para os grupos da tradição Tupiguarani, podem ser citados os trabalhos de Lathrap (1970), Brochado (1973, 1984), Meggers & Evans (1973), Susnik, (1975), Noelli (1993), Monticelli (1995), Landa (1995), Assis (1996), Soares (1997); Moraes (2007), Correa (2014) e Bonomo *et al* (2015). De maneira geral, esses estudos são abrangentes e propõem hipóteses sobre a chegada e dispersão desses grupos ceramistas no sul e sudeste do Brasil, não havendo muita ênfase no papel do estado de São Paulo nesse processo.

Além dos autores supracitados, outros enfatizaram o papel da região atualmente conhecida como o estado de São Paulo na dispersão desses grupos ceramistas. Afonso (2005) destaca que a região entre os vales dos rios Tietê e Paranapanema seriam fundamentais para ocupação Kaingang, e Rodrigues (2007) elabora estudos nos quais são reforçadas essas temáticas, principalmente nas regiões do Rio Feio e Aguapeí (oeste do estado de São Paulo), colocando tais locais como basais para a identidade étnica e pertencimento à etnia.

Araujo (2007) faz uma compilação a respeito dos últimos 35 anos de discussão sobre o grupo Jê e sugere três opções para a nominação dessa tradição: a) utilizar o termo Tradição Eldoradense, uma vez que a mesma foi reconhecida dez anos antes na Argentina, e a precedência do termo é inquestionável; b) utilizar o termo Tradição Itararé, uma vez que a mesma foi definida antes da Tradição Taquara no território brasileiro ou c) utilizar o nome composto Tradição Itararé-Taquara. Neste artigo optamos pela ultima opção.

Embora exista uma volumosa bibliografia a respeito dos Kaingang, os estudos que indicam como se deu a chegada desses grupos ceramistas nas regiões sul e sudeste do Brasil são poucos, afirmam Mota & Assis (2008) e Mota (2016). Essa ideia da dispersão dos grupos Jê é retomada por Pinheiro (1999), Mota (1998, 2000, 2016), Mota & Noelli (1999), Rodrigues (2007) e Schmitz (2007), que indicam a chegada desses ceramistas na região sul do Brasil oriundos dos planaltos entre as nascentes do rio São Francisco e Araguaia no Brasil central. Apesar da dificuldade em definir um momento exato da chegada desses grupos à região, os pesquisadores apontam a existência de diversos registros da ocupação dos grupos Jê no local.

Mageste (2017) relata que a partir dos anos de 1980 e 1990 existe um movimento crescente na arqueologia brasileira que se compromete com a temática de análises espaciais, tentando diferenciar o que é o 'padrão' do que é o 'sistema' de assentamentos, a partir de referências teóricas do processualismo americano e inglês. O autor afirma ainda que pesquisadores como Paulo De Blasis, Maria Dulce Gaspar e Erika Robrahn-Gonzáles, que realizavam análises em diferentes escalas espaciais na região sul de São Paulo, tiveram como intuito a detecção de padrões nos assentamentos regionais.

Segundo Mageste (2017), a definição conceitual e a articulação em termos práticos consistem em: padrão de assentamento sendo o levantamento dos dados referentes à análise espacial dos sítios e sua implantação na paisagem, e a sistematização dos dados

permitiria o estabelecimento de inferências sobre os sistemas de assentamento desde que prezasse por uma visão mais articulada e sincrônica dos sítios de uma região. O autor ainda relata que, a partir da noção de sistemas socioculturais em agrupamentos de sítios arqueológicos, seria possível identificar diferentes funções para os assentamentos, principalmente quando apresentassem distinções na sua implantação na paisagem e cultura material evidenciada. Mageste (2017) entra para o rol de pesquisadores que prezam estudos que extrapolam a noção de análise do espaço e de contextos de um único sítio para um território em escala regional, tal como os trabalhos de Brochado (1984) e Noelli (1993), que também subsidiaram essa abordagem em escala regional.

ESTATÍSTICA E ARQUEOLOGIA

Para Medri (2011), o termo a estatística pode ser entendido como um conjunto de técnicas para planejar experimentos no intuito de obter dados e organizá-los, resumi-los, analisá-los, interpretá-los e assim, por final, extrair conclusões. Medri (2011) explica que a estatística é dividida em duas áreas principais: a descritiva e a inferencial (indutiva). A estatística descritiva, de modo geral, preocupa-se com a organização, apresentação e sintetização dos dados, já a inferencial tem um conjunto de técnicas que permite realizar generalizações e inferências sobre uma determinada população: determina o número de observações (tamanho da amostra), um esquema para a seleção de unidades observacionais, apresenta cálculos das medidas estatísticas, determinação de confiança das estimativas, significâncias dos testes estatísticos, precisão dessas estimativas, entre outros enfoques (MEDRI, 2011).

Por tratar usualmente de vastas quantidades de dados, a Arqueologia é uma disciplina que naturalmente implica na utilização de estatística descritiva e inferencial, a fim de, respectivamente, descrever de forma geral as tendências dos dados e testar hipóteses acerca de sua distribuição. Autores como Myers (1950) e Spaulding (1953) introduziram o tema na primeira metade do século XX, porém o uso de testes estatísticos inicia-se de forma mais intensa e sistemática nos anos de 1960 (CLARKE, 1968), com o advento da Arqueologia Processual, dada sua missão de tornar a Arqueologia uma disciplina mais científica (para um exemplo, vide BINFORD, 1962). A preocupação com a predição e a modelagem incluíram não apenas dados coletados a partir de artefatos (como medidas contínuas e discretas), mas também questões relacionadas à prospecção de sítios arqueológicos e amostragem de sítios e coleções (COWGILL, 1970, DUNNELL & DANCEY, 1983; REDMAN, 1974).

Essa ênfase na estatística como ferramenta para entender padrões no registro arqueológico também pode ser verificada na Arqueologia Evolutiva, nascida na década de 1970, uma vez que a construção de linhagens de artefatos ao longo do tempo e sua interpretação exigem um entendimento das tendências gerais desses dados (DUNNELL, 1980). Ainda, nos últimos anos, a Arqueologia Evolutiva tem se aproximado de áreas como a morfometria geométrica, que também necessita de grande contribuição da estatística para sua aplicação (OKUMURA & ARAUJO, 2019). Finalmente, o refinamento das datações radiocarbônicas (BUCK et al., 1991), assim como a popularização das análises de isótopos estáveis (PEARSON & GROVE, 2013), também contribuíram para um aumento no interesse de análises de cunho estatístico, incluindo a estatística Bayesiana (OTÁROLA-CASTILLO & TORQUATO, 2018). De modo geral, pode-se dizer que a estatística é uma ferramenta para descrever e analisar dados que poderia, em tese, ser usada por qualquer escola da Arqueologia, uma vez que tais abordagens teóricas é que teriam seu papel fundamental na interpretação dos resultados estatísticos.

O PROGRAMA R E SUA APLICAÇÃO NA ARQUEOLOGIA

O programa R é um ambiente computacional para a estatística e produção de gráficos. É um projeto GNU¹ que é semelhante ao ambiente S² que foi desenvolvido na Bell Laboratories (anteriormente AT&T, agora Lucent Technologies) por John Chambers e colegas. O R pode ser considerado como uma implementação do S e, embora existam diferenças importantes no código de funcionamento do programa R, suas maneiras de funcionamento são semelhantes.

De acordo com os desenvolvedores do programa³, o R fornece uma grande variedade de estatísticas (modelagem linear e não linear, testes estatísticos clássicos, análise de séries temporais, classificação, agrupamento etc), técnicas gráficas e é altamente extensível. O programa também tem como ponto forte a facilidade da utilização da interface. Ainda, é garantido que o acesso aos símbolos matemáticos e fórmulas é rápido e o usuário tem total controle para o design dos padrões dos gráficos.

O site de divulgação do programa informa que o *software* R é disponível como Software Livre sob os termos da Free Software Foundation da GNU General Public License em forma de código fonte. Ele funciona em uma grande variedade de plataformas UNIX e sistemas similares (incluindo Free BSD e Linux), Windows e Mac OS.

O ambiente R é um conjunto integrado de instalações de *software* para manipulação de dados, cálculos e exibições gráficas. Inclui um mecanismo eficaz de armazenamento de dados, um conjunto de operadores para cálculos em *arrays*, que é um conjunto coerente e integrado de ferramentas para análise de dados, em que se elaboram gráficos, e a exibição pode ser em tela ou em cópia impressa dos *layouts*. Os desenvolvedores afirmam ter uma linguagem de programação bem desenvolvida, simples e eficaz.

No site de divulgação do R, é relatado que o *software* é projetado em torno de uma linguagem de computador em que se permitem funcionalidades adicionais definindo novas funções e que grande parte do sistema está escrito no dialeto R, isto é, escolhas algorítmicas prontas e de fácil acesso pelo usuário. Os desenvolvedores relatam que existem cerca de oito pacotes básicos com as bases estatísticas e cálculos, mas que também existem outros de fácil acesso através de sites específicos e que cobrem uma ampla gama de estatísticas modernas.

Apesar das facilidades de utilização do *software* R, ainda existem tímidas aplicações em pesquisas científicas voltadas para a arqueologia. Baxter & Cool (2016:21) afirmam que

archaeological journals such as American Antiquity, Antiquity, Antiquaries Journal, The Archaeological Journal, Britannia, Journal of Roman Archaeology, Medieval Archaeology, Proceedings of the Prehistoric Society etc., to name only some of those on our bookshelves or which we consult regularly. If there is a default in this latter group for data analysis it's probably Excel. If you customize things properly you can get perfectly acceptable graphics with Excel but lots of authors don't (BAXTER & COOL, 2016:21).

Os autores relatam que acompanhando a divulgação científica em periódicos voltados para a temática arqueológica, em grande parte ainda se utiliza a ferramenta do Excel em suas análises matemáticas. A razão presumida para isso, segundo os autores, é que, apesar da gratuidade de softwares como o R, muitos usuários relutam em usá-lo, porque é um programa orientado por comandos e não baseado em menu. Para Baxter e

_

¹ GNU é um sistema operacional para software livre, isto é, respeita a liberdade dos usuários.

² O ambiente S tem referência aos softwares mais populares de estatística, como o SPSS e o MiniTab.

 $^{^3\} https://www.r-project.org/about.html$ Acessado em 25/09/2017 às 15:20.

Cool (2016), esse pode ser o principal complicador para a utilização desse tipo de software estatístico, em que o assunto pode ser visto como um desafio antes mesmo de se envolver com a implementação.

Ben Marwick, arqueólogo e membro da Society for American Archaeology (SAA), realizou workshop⁴ em abril de 2017 para apresentar a utilização desse programa junto à comunidade arqueológica. Além disso, esse pesquisador alimenta um blog para a divulgação científica⁵ e compartilha a sua experiência e os motivos que o levaram à preferência do uso do R em pesquisas arqueológicas. No Brasil, existe uma iniciativa do Prof. Dr. Leandro Surya Carvalho de Oliveira Silva (Universidade Federal do Vale do São Francisco – UNIVASF) de divulgação dos trabalhos que utilizam o Programa R para a arqueologia. Tal pesquisador implementou um minicurso intitulado "Introdução à linguagem de programação R" dentro da III Semana de Arqueologia, I Simpósio de Pós-Graduação em Arqueologia: 10 anos do Bacharelado em Arqueologia na Universidade Federal do Piauí - UFPI, em 2018⁶. Em 2019, o Dr. Alex Martire do Museu de Arqueologia e Etnologia da Universidade de São Paulo disponibilizou *online* um curso introdutório para o R⁷. Martire reforça a falta de bibliografia base para a aplicação do R em Arqueologia.

Marwick (2016) apresenta um artigo com os moldes para o compartilhamento dos dados da pesquisa e também dos seus métodos de análises. Esse autor relata que essa maneira de compartilhamento das informações amplia o impacto e os números de citações dos trabalhos. Em relação à linguagem computacional, ele afirma que o R fornece acesso a mais métodos de análises estatísticas que os pacotes dos *softwares* como o Excel e o SPSS. O R, segundo Marwick (2016), é o mais conhecido pela facilidade em visualização dos dados e construção de mapas com poucas linhas de códigos e que estes podem ser reutilizados em novos projetos.

Nesse sentido, Marwick (2016) relata que os problemas técnicos são resolvidos, baseando-se em uma linguagem de código aberto, e os desafios devem ser superados para que exista a mudança nas normas para a replicação dos métodos e melhoria na qualidade dos trabalhos.

O método de utilização do R em arqueologia é também exposto por Carlson (2017), que apresenta um manual que explicita a utilização desses métodos estatísticos para a arqueologia, sendo um guia prático de uso do sistema de computação estatística R escrito especificamente para arqueólogos. A obra de Carlson (2017) é dividida em três trechos que mostram como usar o sistema para analisar muitos tipos de dados arqueológicos. A primeira parte inclui um tutorial em R, com aplicações em dados arqueológicos reais que mostram como calcular estatísticas descritivas, criar tabelas e produzir uma grande variedade de gráficos. Na segunda parte desse manual são apresentadas as principais abordagens estatísticas multivariadas utilizadas pelos arqueólogos, incluindo a regressão

⁴ https://osf.io/2dfhz/wiki/SAA%202017%20Workshop:%20Using%20R%20for%20Archaeological%20Data%20Analysis %2C%20Mapping%2C%20and%20Visualization/ Acessado em 25/9/2017 (HARRIS, M. Convite para workshop do Programa R da University of Washington – EUA, 2017

 $^{^5 \}text{ https://www.r-bloggers.com/doing-quantitative-archaeology-with-open-source-software/} \ Acessado \ em \ 25/09/2017.$

⁶ Surya ainda orientou um trabalho de iniciação científica (IGUATEMY, Mateus Maurício de Mello. Sítio Ponte Velha: desenvolvimento de aplicações para trabalhos de campo baseados em linguagem R. 2013. Iniciação Científica. (Graduando em Arqueologia e Preservação Patrimonial) - Universidade Federal do Vale do São Francisco, Coordenação de Aperfeiçoamento de Pessoal de Nível Superior). Além disso, existe um artigo do autor (SURYA, L.. A linguagem R e suas possibilidades de uso em Arqueologia. In: XVII Congresso da Sociedade de Arqueologia Brasileira, 2013, Aracaju. Anais do XVII Congresso da Sociedade de Arqueologia Brasileira, 2013 (no Prelo).

⁷ MARTIRE, A. R-chaeology: uma introdução à programação R em Arqueologia. Disponível em: http://www.arise.mae.usp.br/courses/r-chaeology-uma-introducao-a-programacao-r-em-arqueologia/?tab=tab-overview Acessado em 29 de janeiro de 2019.

múltipla (e o modelo linear generalizado); análise múltipla de variância e análise discriminante; análise de componentes principais; análise de correspondência; distâncias e escala; e análise de *cluster* e a terceira parte cobre temas especializados em arqueologia, incluindo análise espacial intra-sítio, seriação e diversidade dos conjuntos. A escassa apresentação de bibliografia aqui exposta e a utilização desse *software* para a arqueologia sugerem um cenário que permite o aprofundamento e a ampliação desses métodos aplicados aos dados arqueológicos.

OS DADOS

Os dados utilizados aqui são provenientes do Banco de dados (BD) elaborado por Perez (2018) e que está disponível para consulta na Biblioteca do Museu de Arqueologia e Etnologia da Universidade de São Paulo (MAE/USP) e reúne subsídios preponderantemente de dois centros de conhecimento de Arqueologia Paulista: Instituto do Patrimônio Histórico e Artístico Nacional, Superintendência de São Paulo – IPHAN/SP e a Biblioteca do Museu de Arqueologia e Etnologia da Universidade de São Paulo.

Neste BD são catalogadas informações provenientes de: 57 relatórios de diversas empresas de arqueologia preventiva, 42 artigos de revistas, 12 teses, 10 dissertações, cinco obras literárias específicas ao tema, duas obras de publicação de acervo, um inventário organizado por José Luiz de Morais em 2001, uma citação classificada como 'outro', o que soma 130 obras, além das fichas cadastrais de sítios arqueológicos, sendo 132 fichas disponíveis online no site do Cadastro Nacional de Sítios Arqueológicos (CNSA) e outras 12 caixas-arquivos contendo as fichas impressas para o Estado de São Paulo todo – esses dados correspondem ao que foi produzido desde o ano de 1895 (Revista do Museu Paulista) até o ano de 2015, quando se encerraram as pesquisas junto ao acervo do IPHAN/SP.

Em uma visão geral, o BD conta com 783 sítios cadastrados, sendo que são 482 sítios da Tradição Tupiguarani, 195 sítios da Tradição Itararé-Taquara, 97 sítios sem referência a uma Tradição específica e 9 sítios que são classificados como pertencendo às duas tradições Tupiguarani e Itararé-Taquara, sendo um deles descrito como da Tradição Aratu-Sapucaí (Sítio Chopim no município de Castilho – SP, identificado durante os trabalhos do Projeto de Salvamento Porto Primavera em 19978). Esses sítios arqueológicos estão distribuídos por 121 municípios, correspondendo a 41,6% do Estado de São Paulo. Ainda segundo as informações contidas no BD existem 77 sítios que são datados (9,8%); 418 sítios com coordenadas geográficas conhecidas (53,3%); e apenas 44 sítios, ou 5,6% deles, com ambas as características (PEREZ, *et al.*, 2018a).

OS RESULTADOS ALCANÇADOS

Neste artigo são apresentadas informações necessárias para a compreensão da caracterização e diferenciação geral dos sítios arqueológicos entre os grupos ceramistas das tradições Tupiguarani e Itararé-Taquara no oeste e sul do Estado de São Paulo. Os dados inseridos neste item referente à analise descritiva são de fundamental importância para a compreensão da análise inferencial mais à frente, pois nesta descrição foram oferecidos apenas atributos que apresentam importância estatística para a observação de diferenças entre os grupos ceramistas, sendo esse o objetivo da análise. Ainda, destaca-se

-

⁸ IV Relatório Científico Relativo ao Projeto de Salvamento Arqueológico de Porto Primavera - SP, de acordo com autorização constante na Portaria nº. 38, de 1º de setembro de 1997, do IPHAN.

que tais dados foram agrupados por frequência com o uso do software Excel da Microsoft, com base nos dados do BD, anteriormente citado.

Medri (2011) destaca que a utilização desse tipo de cálculo (frequências) é importante quando existe uma grande quantidade de atributos iguais, e a escolha dos intervalos é sempre arbitrária e cabe ao pesquisador escolhê-las. Durante a construção das frequências, determinar um pequeno número de classes pode levar à perda de informações, enquanto ao utilizar grandes quantidades de classes se prejudica o objetivo de análise resumida. Esses dois extremos variam de acordo com grau de suavização da representação gráfica dos dados escolhidos pelo pesquisador.

A escolha desse método se justifica pelo tipo de dado analisado nesta pesquisa, que pode ser resumido como sendo qualitativo, com valores faltantes e amostras numericamente desiguais⁹. Ainda lembramos que nesta pesquisa tivemos o auxílio de uma equipe de professores e estudantes do Departamento de Estatística do Instituto de Matemática e Estatística da Universidade de São Paulo, através do Centro de Consultoria em Estatística Aplicada (CEA), que foi responsável pelo manejo dos dados dentro do Programa R.

ANÁLISE DESCRITIVA - CATEGORIZAÇÃO DOS SÍTIOS ESTUDADOS

Neste item são expostos os resultados da análise descritiva realizada a partir do sequenciamento das frequências dos atributos juntos ao Banco de Dados (BD) realizado por Perez (2018), em que este método teve por intuito identificar as características espaciais típicas para cada Tradição arqueológica, sintetizando a distinção entre os grupos. Perez (2018) aponta que os sítios da <u>Tradição Tupiguarani</u> estão localizados mais ao oeste da região estudada, que é compreendida pelo oeste e sul paulistas. Segundo o autor, as características que mais diferenciam os sítios desta Tradição dos sítios da Tradição Itararé-Taquara são:

- Deposição: superfície e profundidade 54,9% (130);
- Compartimento topográfico: planície de inundação 49,4% (85);
- Unidade geomorfológica: planície 45,5% (96);
- Dimensões: maiores;
- Altimetria (Cota): menores em relação ao nível do mar (menores que 350m).

Em relação às variáveis referentes à vegetação, Perez (2018) destaca que a maior diferença foi observada em:

- Vegetação: Floresta Estacional Semidecidual 87,5% (124);
- Geomorfologia: Planalto Centro Ocidental 54,3% (118);
- Declividade média do terreno: menores declividades;
- Classe de rocha: Sedimentar 70,4% (153);

Por fim, Perez (2018) expõe que para as variáveis relativas ao clima, tem-se:

- Tipo de Seca: 1 a 2 meses secos 54,4% (118);
- Temperatura: subquente 52,1% (112);

⁹ Essas características impedem a aplicação, por exemplo, de testes estatísticos paramétricos, como teste t de Student, Q-quadrado e ANOVA. Nenhum desses pode ser utilizado, pois o teste t e Anova exigem a normalidade dos dados quantitativos, e o teste de Q-quadrado não aceita nenhuma contagem como sendo zero.

• Tipo de clima: úmido - 60,4% (131).

Tais números, para a Tradição Tupiguarani, corresponde ao total de 482 sítios. Já para a <u>Tradição Itararé-Taquara</u>, destaca-se por serem os sítios localizados em geral, mais ao sul do Estado de São Paulo e suas características seguem as seguintes classificações (PEREZ, 2018):

- Deposição: superfície 73,2% (131);
- Compartimento topográfico: média vertente 34,1% (41);
- Unidade geomorfológica: Planalto 70,3% (55);
- Dimensões: menores;
- Altimetria (Cota): maiores em relação ao nível do mar (maiores que 450m).

Em relação às variáveis referentes à vegetação, Perez (2018) afirma que a maior diferença foi observada em:

- Vegetação: Floresta Ombrófila Mista 63,5% (40);
- Geomorfologia: Planaltos 67,8% (48);
- Declividade média do terreno: maiores declividades;
- Classe de rocha: Ígnea e metamórfica 33,8% (23);

A descrição para tipo de rochas apresentada por Perez (2018) não é precisa devida à baixa representatividade estatística desses sítios com descrição – são apenas 23 sítios em áreas de rochas ígneas e 26 em áreas de sedimentar – o que não indica diferença estatística para este item.

Por fim, para as variáveis relativas ao clima, as que se destacam são:

- Tipo de Seca: subseca 61,9% (44);
- Temperatura: mesotérmico brando 70,3% (50);
- Tipo de clima: super-úmido 78,8% (56).

Tais dados apresentados aqui para a Tradição Itararé-Taquara tem como total 195 sítios.

De acordo com Esteves *et al.* (2017) e Perez (2018), os dados de altimetria (cota), vegetação, geomorfologia, rochas e variáveis climáticas são fundamentais para a análise inferencial e os modelos preditivos da ocupação regional pelos grupos ceramistas e, por esse motivo, serão descriminados extensivamente. Além disso, Esteves *et al.* (2017) explicam que com essa categorização é possível identificar a Tradição arqueológica a que se filia um sítio observando esses atributos descritivos, oportunizando assim a identificação de uma Tradição ceramista aos grupos de sítios que não apresentam tal categoria. Nesse sentido, seguem os dados das categorias utilizadas.

ALTIMETRIA (COTA) - ALTITUDE EM RELAÇÃO AO NÍVEL DO MAR

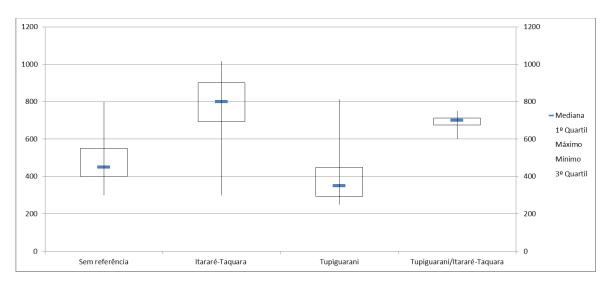
Dos 783 sítios considerados no estudo, 345 foram observados em relação à altimetria (cota), que corresponde à altitude em relação ao nível do mar em que se encontram os sítios arqueológicos. A Tabela 1, juntamente com o Gráfico Boxplot 1, indica que os sítios da Tradição Itararé-Taquara geralmente estão em regiões com uma altitude mais elevada, uma vez que a sua mediana é a maior (800m), quando comparada aos outros

grupos de Tradição Tupiguarani que apresentam a menor mediana (350), ou seja, os grupos com essa Tradição habitam em regiões com altitude menos elevadas.

Tabela 1 - Tabela de Frequências de Altimetria (Cota) por Tradição. Fonte: Perez (2018).

Altimetria (Cota)	Sem referência	Itararé- Taquara	Tupiguarani	Tupiguarani/Itararé- Taquara
Mínimo	300	300	250	600
1º Quartil	400	693,8	293,8	675
Mediana	450	800	350	700
Média	497,4	741	389,4	687,5
3º Quartil	550	900	450	712,5
Máximo	800	1016,5	812,9	750
Dados faltantes	44	124	265	5

Gráfico 1 - Boxplot para a variável Altimetria (Cota) por Tradição.



VEGETAÇÃO

O atributo referente à *vegetação* do espaço foi atribuído a 244 dos 783 sítios do estudo. Essa categoria corresponde à vegetação atual encontrado no Estado de São Paulo. De acordo com Saia (2006), o clima e precipitação pretéritos são semelhantes aos atuais na região do Vale do rio Ribeira do Iguape, o que possibilita inferir que a vegetação do período da ocupação dos grupos ceramistas é semelhante ao atual na região, embora seja complexo estender as avaliações de Saia (2004) para o extremo oeste paulista. Pela Tabela 2, a característica Floresta Estacional Semidecidual está presente em 87,5% dos sítios com Tradição Tupiguarani. Já a característica Floresta Ombrófila Mista está em 63,5% dos sítios com Tradição Itararé-Taquara.

Tabela 2 - Tabela de Frequências da Vegetação pretérita (porcentagem e contagem) do espaço por Tradição. Fonte: Perez (2018).

Vacatasão	Sem	Itararé-	Tupiguarani	Tupiguarani/Itararé-
Vegetação	referência	referência Taquara		Taquara
Floresta				
Estacional	65,5% (23)	20,7% (13)	87,5% (124)	0,0% (0)
Semidecidual				
Floresta	5,8% (2,1)	0,0% (0)	1,4% (2)	0,0% (0)
Ombrófila Densa	3,670 (2,1)	0,070 (0)	1,470 (2)	0,0% (0)
Floresta	14,4% (5)	63,5% (40)	1,4% (2)	0,0% (0)
Ombrófila Mista	14,470 (3)	03,370 (40)	1,4/0 (2)	0,0% (0)
Savana	14,4% (5)	15,8% (10)	9,8% (14)	100,0% (4)
Total	100% (35)	100% (63)	100% (142)	100% (4)

Na Tabela 2 percebe-se a preponderância dos grupos Tupiguarani em Floresta Estacional Semidecidual e para a Tradição Itararé-Taquara, Floresta Ombrófila mista, e não indica claramente alguma semelhança do comportamento dos sítios sem referência com alguma Tradição em relação à vegetação pretérita do espaço.

GEOMORFOLOGIA

Dos 783 sítios considerados no estudo, 344 foram classificados segundo sua geomorfologia – dados elencados a partir do SIG e com precisão das coordenadas geográficas. A Tabela3 mostra que a maior parte dos sítios sem referência (66%) e com Tradição Tupiguarani (54,3%) tem como relevo Planalto Centro Ocidental Indiferenciado, enquanto 67,8% dos sítios com Tradição Itararé-Taquara tem como relevo Planalto. Foram considerados apenas 4 sítios com ambas as tradições, o que não permite uma análise mais profunda sobre eles.

Tabela 3 - Tabela de Frequências da Geomorfologia (porcentagem e contagem) por Tradição. Fonte: Perez (2018).

Geomorfologia	Sem referência	Itararé- Taquara	Tupiguarani	Tupiguarani/Itararé- Taquara
Corpos d'água	5,7% (3)	1,4% (1)	18,0% (39)	0,0% (0)
Depressão indiferenciada	17,0% (9)	2,8% (2)	10,6% (23)	50,0% (2)
Planalto centro ocidental indiferenciado	66,0% (35)	25,3% (18)	54,3% (118)	0,0% (0)
Planaltos	11,3% (6)	67,8% (48)	3,8% (8)	50,0% (2)
Planícies fluviais	0,0% (0)	1,4% (1)	13,3% (29)	0,0% (0)
Serras/escarpas	0,0% (0)	1,4% (1)	0,0% (0)	0,0% (0)
Total	100% (53)	100% (70)	100% (217)	100% (4)

Comparando as informações das unidades geomorfológicas apresentadas pela revisão bibliográfica e as elencadas pelo SIG, nota-se a diferenciação entre as nominações dadas às regiões de ocupação Tupiguarani. Enquanto na bibliografia têm-se áreas de planície, o SIG indica áreas do Planalto Centro Ocidental. Para os sítios de Tradição Itararé-Taquara, as nominações são as mesmas, isto é, planaltos. Os sítios sem atribuição a uma tradição arqueológica estão entre Planaltos e Planícies, para os dados do BD e nos dados do SIG, são considerados planaltos.

CLASSE DE ROCHA

Para classe de rocha, 340 dos 783 sítios considerados no estudo foram classificados. Na Tabela 4, a classe mais presente foi a sedimentar: 70,4% dos sítios com Tradição Tupiguarani e 33,8% dos sítios com Tradição Itararé-Taquara apresentam a característica.

Tabela 4 - Tabela de Frequências de Classe de rocha (porcentagem e contagem)por Tradição. Fonte: Perez (2018).

Classe de rocha	Sem ref	erência	Itararé-	Гаquara	Tupig	uarani	Tupigu Itararé-T	
Ígnea	30,2%	(16)	17,8%	(12)	29,1%	(63)	50%	(1)
Ígnea, Metamórfica	3,8%	(2)	33,8%	(23)	0,0%	(0)	0%	(0)
Metamórfica	1,8%	(1)	10,3%	(7)	0,4%	(1)	0%	(0)
Sedimentar	64,2%	(34)	38,1%	(26)	70,4%	(153)	50%	(1)
Total	100%	(53)	100%	(68)	100%	(217)	100%	(2)

VARIÁVEIS DE CLIMA

Para as três variáveis deste grupo, foram classificados 344 dos 783 sítios considerados no estudo. Vale ressaltar que há apenas 4 sítios com ambas tradições classificados nestas variáveis, o que não permite uma análise mais profunda dessas observações.

TIPO DE SECA

A Tabela 5 mostra que os sítios sem referência e com Tradição Tupiguarani estão mais distribuídos do que os sítios com Tradição Itararé-Taquara:

- 54% dos sítios Tupiguarani foram classificados como 1 a 2 *meses secos* e 40% como *subseca*;
- 20,7% dos sítios sem referência foram classificados como 1 a 2 meses secos, 43,4% como subseca e 35,9% como 3 meses secos;
- 62% dos sítios com Tradição Itararé-Taquara foram classificados como subseca.

É importante notar que 216 sítios com Tradição Tupiguarani foram analisados nesta variável, sendo 53 sítios sem referência e 71 com Tradição Itararé-Taquara.

REVISTA DE ARQUEOLOGÍA VOLUME 32 No. 1 2019: 84-107

T: 1	Sem	Itararé-	T:	Tupiguarani/Itararé-	
Tipo de seca	referência	Taquara	Tupiguarani	Taquara	
1 a 2 meses secos	20,7% (11)	4,1% (3)	54,4% (118)	0,0% (0)	
3 meses secos	35,9% (19)	17,0% (12)	6,0% (13)	0,0% (0)	
Sem seca	0,0% (0)	17,0% (12)	0,0% (0)	0,0% (0)	
Subseca	43,4% (23)	61,9% (44)	39,6% (85)	100,0% (4)	
Total	100% (53)	100% (71)	100% (216)	100.0% (4)	

Tabela 5 - Tabela de Frequências do Tipo de seca (porcentagem e contagem) por Tradição. Fonte: Perez (2018).

Na Tabela 5 é mostrado que não é possível notar claramente uma semelhança no comportamento dos sítios sem referência com alguma tradição.

TEMPERATURA

Pela Tabela 6, observa-se que 70,3% dos sítios Itararé-Taquara foram classificados como *mesotérmica branda*, porém aqueles com Tradição Tupiguarani estão mais distribuídos: 52,1% deles foram caracterizados com uma temperatura *subquente* e 45,2% como *quente*. A maioria dos sítios sem referência (67,9%) foi classificada como *subquente*. Assim, não é possível notar uma semelhança no comportamento dos sítios sem referência com os sítios de alguma tradição.

Tabela 6 - Tabela de Frequências de Temperatura (porcentagem e contagem) por Tradição. Fonte: Perez (2018).

Temperatura	Sem referência	Itararé- Taquara	Tupiguarani	Tupiguarani/Itararé- Taquara
Mesotérmica branda	13,2% (7)	70,3% (50)	2,7% (5)	75,0% (3)
Quente	18,9% (10)	1,4% (1)	45,2% (97)	0,0% (0)
Subquente	67,9% (36)	28,3% (20)	52,1% (112)	25,0% (1)
Total	100% (53)	100% (71)	100% (216)	100% (4)

Diante do que foi apresentado passamos a entender a análise inferencial e a execução dos modelos a partir da utilização do *software* R e os resultados alcançados com o uso dele.

ANÁLISE INFERENCIAL - MODELOS DE REGRESSÃO LOGÍSTICA

Nessa análise, o objetivo é indicar, a partir da construção de um modelo de regressão logística, as características geoespaciais mais presentes em cada tradição arqueológica. A regressão logística é uma técnica estatística que tem como objetivo produzir um conjunto de observações de modelo que permita a predição de valores tomados por uma variável categórica, a partir de uma série de variáveis explicativas contínuas e/ou binárias. Em comparação com as técnicas conhecidas em regressão, em especial a regressão linear, a regressão logística distingue-se essencialmente pelo fato de a variável ter uma resposta categórica. A regressão linear é uma equação para se estimar a condicional (valor esperado) de uma variável y, dados os valores de algumas outras

variáveis x. A regressão, em geral, tem como objetivo tratar de um valor que não se consegue estimar inicialmente.

Para a estimação do modelo de regressão, foram considerados apenas os sítios arqueológicos que continham em suas descrições todas as variáveis (altimetria, vegetação, geomorfologia, declividade média, classe de rocha e para o clima: período de seca). No total, foram utilizados 214 sítios Tupiguarani e 60 sítios Itararé-Taquara. Além disso, foi criada uma variável que identificou a tradição arqueológica do sítio mais próximo dele. Essa proximidade foi calculada através da distância euclidiana entre sítios. A Distância Euclidiana (DE) é a distância entre dois pontos que pode ser provada pela aplicação repetida do teorema de Pitágoras.

Calculada essa distância euclidiana, obteve-se o sítio mais próximo para cada um dos sítios e suas tradições arqueológicas correspondentes. Em caso de empate na distância euclidiana, foi considerada a tradição arqueológica mais frequente entre os sítios obtidos.

MODELO DE REGRESSÃO LOGÍSTICO CLÁSSICO

As variáveis selecionadas para compor os modelos de regressão logística apresentados neste artigo foram: altimetria, vegetação geomorfologia, declividade média, classe de rocha, seca, temperatura e o sítio mais próximo elaborado a partir do cálculo da Distância Euclidiana (DE).

Na Tabela7 são apresentadas as características que serviram de intercepto (ou modelo que é o coeficiente que irá guiar a modelagem) para a modelagem, sendo as características descritas em cada variável e o seu respectivo nível.

Variável	Valor/Nível
Altimetria	0
Vegetação Atual	Florestamento/Reflorestamento com eucaliptos
Vegetação Pretérita	Floresta ombrófila densa
Geomorfologia	Serras/escarpas
Classe de Rocha	Metamórfica
Seca	0 a 2 meses secos
Temperatura	Mesotérmico brando
Sítio mais próximo	Itararé-Taquara

Tabela 7 - Componentes do intercepto. Fonte: Perez (2018).

Definido o intercepto (modelo), foi estimado um modelo considerando todas as variáveis elencadas anteriormente. As variáveis que não apresentavam significância estatística foram excluídas por etapas, fixando o nível de significância do teste em 10% de acordo com seu valor de p. Assim, o modelo final obtido está listado na Tabela 8, juntamente com as estimativas dos coeficientes, o erro-padrão, o valor de p e a razão de chances (OR) e seus respectivos limites inferior (OR_i) e superior (OR_s). Lembrando que OR relaciona-se ao valor da Curva ROC e que, segundo Esteves et al. (2017), a Curva ROC é uma representação gráfica que ilustra o desempenho (ou performance) de um sistema classificador binário e como o seu limiar de discriminação é variado, ela é obtida pela representação da fração entre Positivos Verdadeiros e Positivos Totais (RPV=PV/P) versus a fração entre Positivos Falsos e Negativos Totais (RPF=PF/N), em várias configurações de limites. Esteves et al. (2017) relatam que a análise de Curva ROC

possibilita selecionar modelos ótimos e descartar os de qualidade inferior a partir do contexto de custo ou da distribuição de uma variável. As medidas fornecidas pela Curva ROC apresentam a proporção como o modelo classifica um sítio com Tradição Tupiguarani e um sítio com Tradição Itararé-Taquara, afirmam Esteves *et al.* (2017). A Curva ROC é a representação dos eixos (1-Especificidade, Sensibilidade) e a área debaixo da curva representa uma medida de desempenho do modelo de classificação.

Variável	Estimativa do Coeficiente	Erro Padrão	OR_i	OR	ORs	p
Intercepto	7,114	1,705	-	-	-	<0,001
Altimetria (Cota)	-0,014	0,003	0,980	0,986	0,992	<0,001
Seca – 3 meses secos	-2,109	-0,922	0,739	0,121	0,0199	0,022
Seca – Subseca	2,643	1,163	1,438	14,055	137,343	0,022
Sítio mais próximo – Tupiguarani	2,242	0,649	2,638	9,412	33,584	<0,001
Sítio mais próximo – Tupiguarani/Itararé- Taquara	-1,863	1,064	0,019	0,155	1,249	0,080

Tabela 8 - Modelo Clássico Final. Fonte: Perez (2018).

Cada coeficiente do modelo é interpretado individualmente para entender a influência de cada variável na classificação do sítio segundo Tradição arqueológica. Assim, quando uma variável é analisada, diz-se que as variáveis restantes são mantidas constantes. Nesse sentido, a interpretação de cada coeficiente do modelo é relacionada ao intercepto (modelo) referência, descrita a seguir.

Altimetria (Cota)

A estimativa do coeficiente de altimetria (Cota) é igual a -0,014. Como $e^{-0,014}$ = 0,985, então com o acréscimo de 1 unidade de Altimetria (Cota), a chance do sítio ser classificado como Tradição Tupiguarani é multiplicado por 0,985.

Seca – 3 meses secos

A estimativa do coeficiente de Seca – 3 meses secos é igual a -2,109. Como $e^{-2,109} = 0,121$, então a chance do sítio com 3 meses secos ser classificado como Tradição Tupiguarani é 0,121 vezes a chance do sítio com 0 a 2 meses secos ser classificado do mesmo modo.

Seca - Subseca

A estimativa do coeficiente Seca – Subseca é igual a 2,643. Como $e^{2,643} = 14,059$, então a chance do sítio com subseca ser classificado como Tradição Tupiguarani é aproximadamente 14 vezes a chance do sítio com 0 a 2 meses secos.

• Sítio mais próximo - Tupiguarani

A estimativa do coeficiente de Sítio mais próximo, Tupiguarani, é igual a 2,242. Como $e^{2,242} = 9,412$, então a chance do sítio cujo sítio mais próximo tem Tradição

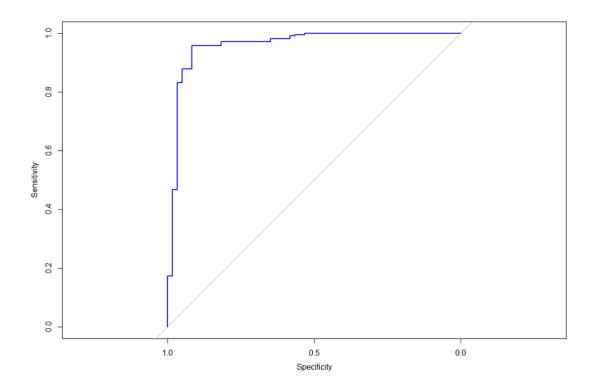
Tupiguarani ser classificado como Tradição Tupiguarani é 9,412 vezes a chance do sítio cujo sítio mais próximo tem Tradição Itararé-Taquara.

• Sítio mais próximo - Tupiguarani/Itararé-Taquara

A estimativa do coeficiente de Sítio mais próximo – Tupiguarani/Itararé-Taquara é igual a -1,863. Como $e^{-1,863} = 0,155$, então a chance do sítio cujo sítio mais próximo tem ambas as tradições ser classificado como Tradição Tupiguarani é 0,155 vezes a chance do sítio cujo sítio mais próximo tem Tradição Itararé-Taquara.

Como forma de analisar a qualidade do modelo de Regressão Clássico, foi construída a Curva ROC, representada no Gráfico 2. A partir dele, foi definido o ponto de corte para a probabilidade de classificação do sítio como Tupiguarani, de forma a classificar a Tradição com maior precisão possível. Para o modelo de Regressão Clássico, esse ponto de corte foi 0,5, o que resulta em uma probabilidade de acerto de aproximadamente 95%.





Além disso, o programa R permite a elaboração de uma figura (Figura 1) que representa a probabilidade de um sítio ser classificado em uma escala de cores (do rosa ao azul) e símbolos (triângulos ou círculos). Nessa classificação, a figura apresenta sítios tipicamente Tupiguarani, com símbolos de triângulos rosa, e os sítios tipicamente Itararé-Taquara, com símbolos de círculos azuis. O modelo de regressão Clássico permite a avaliação dos sítios arqueológicos segundo o intercepto apresentado anteriormente. O modelo parte do princípio de que os sítios sejam Tupiguarani e calcula o distanciamento do intercepto proposto nesta análise. O resultado alcançado sugere que grande parte deles são atribuídas às classificações: Tupiguarani (triângulo), como Tupiguarani (rosa), e Itararé-Taquara (círculo), como Itararé-Taquara (azul). Apesar disso, alguns sítios são colocados com classificações inversas (triângulos azuis e círculo

rosa). São nesses locais de inversão das classificações que se pode perceber possíveis questões de fronteira ou áreas de contato entre os grupos ceramistas. Nessa figura apresentada é possível observar a plotagem dos sítios arqueológicos na região estudada, mesmo com uma distorção¹⁰ referente à disposição deles na paisagem.

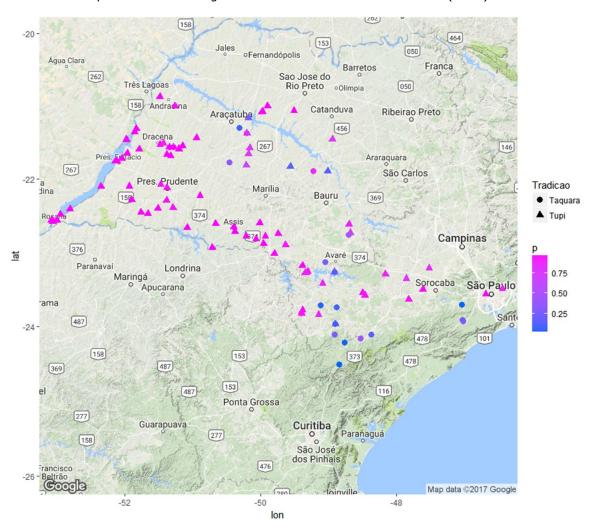


Figura 1 - Mapa de probabilidade de classificação dos sítios segundo Tradição arqueológica pelo método de Regressão Clássico. Fonte: Esteves *et al.* (2017).

MODELO DE REGRESSÃO LOGÍSTICO BAYESIANO

Para contrapor o modelo de Regressão Clássico, foi possível realizar a modelagem de Regressão Bayesiana. De acordo com Esteves *et al.* (2017), tal modelo é inspirado no Teorema de Bayes, que mostra como alterar as probabilidades *a priori*, tendo em vista novas evidências para obter probabilidades a *posteriori*, isto é, o Teorema expressa como a probabilidade de um evento (ou o grau de crença na ocorrência deste evento) deve ser alterada após considerar evidências sobre a ocorrência do evento. Além disso, essa fórmula da Teoria de Bayes, relatam Esteves *et al.* (2017), possibilita a realização de cálculos para a *inferência bayesiana* que é um tipo de inferência estatística que descreve as incertezas sobre quantidades invisíveis de forma probabilística. As incertezas podem ser modificadas periodicamente após as observações dos novos dados ou resultados.

¹⁰ Tal distorção dos pontos apresentados na figura pode ter sido resultado da diferença de escala ou base geográfica utilizada pela equipe de Esteves et al. (2017).

Nessa modelagem foram consideradas as mesmas variáveis explicativas que no modelo de Regressão Clássico Final, com a finalidade de deixar os modelos comparáveis, assim como as mesmas características no intercepto (modelo) – Tabela 9.

Variável	Estimativa	Erro	LB	UB	ORi	OR	ORs
Variavei	Pontual	Padrão	22	CD	Old	OR	Ole
Intercepto	5,171	2,469	0,007	8,889	-	-	-
Altimetria (Cota)	-0,010	0,004	-0,017	-0,002	0,983	0,990	0,998
Seca – 3 meses secos	-1,909	1,006	-3,805	0,063	0,022	0,148	1,065
Seca – Subseca	1,684	1,168	-0,411	3,899	0,663	5,387	49,353
Sítio mais próximo –	2,737	0.781	-1,307	4,359	0,271	15,441	78,179
Tupiguarani	2,,	3,7 5 1	1,007	.,557	0,271	10,111	7 5,17 7
Sítio mais próximo -							
Tupiguarani/Itararé-	-1,758	1,111	-3,965	0,227	0,019	0,172	1,255
Taquara							

Tabela 9 - Modelo Bayesiano Final. Fonte: Perez (2018).

Da mesma forma que no modelo de Regressão Clássico, a interpretação de cada coeficiente é feita considerando o intercepto (modelo) como referência e está descrita a seguir.

• Altimetria (Cota)

A estimativa do coeficiente de Altimetria (Cota) é igual a -0,010. Como $e^{-0,010} = 0,990$, então a chance do sítio ser classificado como Tradição Tupiguarani multiplicado por 0,990, para o aumento de cada unidade de Altimetria (Cota).

• Seca - 3 meses secos

A estimativa do coeficiente de Seca – 3 meses secos é igual a -1,909. Como $e^{-1,909}$ = 0,148, então a chance do sítio com 3 meses secos ser classificado como Tradição Tupiguarani é 0,148 vezes a chance do sítio com 0 a 2 meses secos.

Seca - Subseca

A estimativa do coeficiente Seca – Subseca é igual a 1,684. Como $e^{1,684} = 5,387$, então a chance do sítio com subseca ser classificado como Tradição Tupiguarani é 5,387 vezes a chance do sítio com 0 a 2 meses secos.

Sítio mais próximo – Itararé-Taquara

A estimativa do coeficiente de Sítio mais próximo – Tupiguarani é igual a 2,737. Como $e^{2,737} = 15,440$, então a chance do sítio cujo sítio mais próximo tem Tradição Itararé-Taquara ser classificado como Tradição Tupiguarani é aproximadamente 15 vezes a chance do sítio cujo sítio mais próximo tem Tradição Tupiguarani.

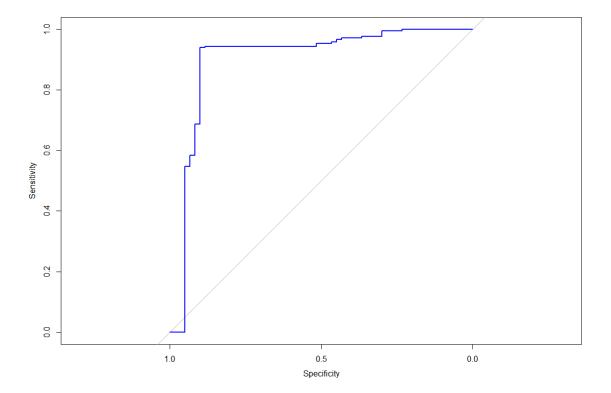
Sítio mais próximo – Tupiguarani/Itararé-Taquara

A estimativa do coeficiente de Sítio mais próximo – Tupiguarani/Itararé-Taquara é igual a -1,758. Como $e^{-1,758} = 0,172$, então a chance do sítio cujo sítio mais próximo

tem ambas as tradições ser classificado como Tradição Tupiguarani é aproximadamente 0,172 vezes a chance do sítio cujo sítio mais próximo tem Tradição Itararé-Taquara.

Assim como feito para o modelo clássico, usamos a Curva ROC (Gráfico 3) para definir o ponto de corte, que, nesse caso, foi 0,44, o que resultou em um acerto de aproximadamente 93%.

Gráfico 3 – Curva ROC das previsões da Regressão Logística Bayesiana. Fonte: Esteves *et al.* (2017).



A Figura 2 representa a probabilidade de classificação segundo tradição arqueológica. Do mesmo modo que foi interpretada a figura 1, a escala de cores estabelece a probabilidade de um sítio ser atrelado a uma ou outra tradição arqueológica, enquanto o símbolo indica se este sítio é Tupiguarani (triângulo) ou Itararé-Taquara (círculo) e Tupiguarani (rosa) e Itararé-Taquara (azul). Nessa figura 2 é possível observar novamente a mesma distorção em relação à localização dos sítios, como aconteceu anteriormente na Figura 1. Alguns sítios também são colocados com classificações inversas (triângulo azul e círculo rosa). Reforça-se, desta maneira, que podem ser nesses locais de inversão de classificações que podem estar inseridos os locais de fronteira entre os grupos estudados.

REUISTA DE ARQUEOLOGIA VOLUME 32 No. 1 2019; 84-107

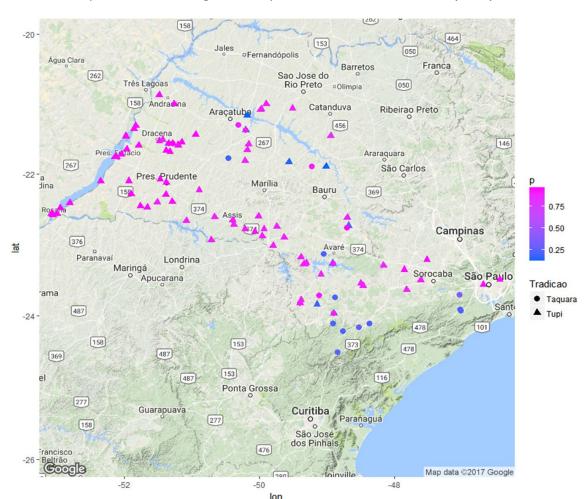


Figura 2 – Mapa de probabilidade de classificação dos sítios segundo Tradição arqueológica pelo método de Regressão Bayesiano. Fonte: Esteves *et al.* (2017).

Tabela 10 - Comparação entre os modelos Clássico e Bayesiano. Fonte: Perez (2018).

Tradição	Tradição arqueológica estimada pelo		Tradição arqueológica estimada pelo		
arqueológica	método	Clássico	método Ba	nyesiano	
observada	Tupiguarani	Itararé-Taquara	Tupiguarani	Itararé-Taquara	
Tupiguarani	96%	4%	94%	6%	
Itararé- Taquara	8%	92%	11%	89%	

Temos que, para ambos os modelos, todas as variáveis estão no mesmo sentido de resposta, isto é, quando o exponencial do coeficiente de um modelo é maior que 1, para o outro modelo, esse exponencial também é maior que 1. Além disso, temos que o acerto foi igual a 95% para o modelo Clássico, enquanto para o modelo Bayesiano foi de 93%.

Nesse sentido, os modelos de regressão Clássico ou Bayesiano poderiam contribuir para inferência satisfatória para os sítios que não foram classificados quanto à sua tradição, tanto na bibliografia consultada quanto para outros sítios que estão alocados na área estudada e que não vieram a ser estudados nesta análise. Apenas a averiguação *in situ* ou mesmo uma comparação mais aprofundada com modelos preditivos propostos poderiam vir a confirmar tais classificações para a tradição arqueológica.

Reforça-se ainda que o estabelecimento de um intercepto (modelo) utilizado para comparação aos modelos de regressão Clássico e Bayesiano foi baseado na análise descritiva dos dados dos sítios arqueológicos, e esses dados foram reproduzidos a partir do Banco de Dados (BD) e do Sistema de Informação Geográfico existente para os sítios dos grupos cerâmicos das regiões oeste e sul do Estado de São Paulo, e que, por esse motivo, esses dados podem apresentar ruído gerado pelo foco específico do estudo em áreas específicas dentro do Estado de São Paulo, isto é, áreas de interesse comercial ou industrial, ou mesmo áreas de hidroelétricas, em que os estudos de arqueologia foram mais realizados. Nesse sentido, o ruído gerado poderia causar alguma outra forma de distorção na análise comparativa dos modelos de regressão, mas esse apenas pode ser corrigido com a amplificação dos estudos de arqueologia regional. Além disso, através da utilização do R na arqueologia, se oportunizou a diferenciação estatística entre os grupos ceramistas, bem como a identificação de possíveis áreas de contato entre eles.

CONSIDERAÇÕES FINAIS

A contribuição deste artigo se deu a partir da utilização de um método até então desconhecido nos estudos de cultura material, mas que aqui conseguiu atingir um objetivo diferente quando tornou possíveis a identificação e a classificação de um sítio arqueológico segundo suas características físicas e espaciais, em que o sítio arqueológico foi tratado como artefato, assim como Gaspar (1991, 1996) conseguia observar os sambaquis, objeto de suas pesquisas – o que deve ser observado também enquanto artefato a ser estudado.

Dessa forma, foi apresentada uma proposta de categorização espacial desses grupos a partir da distribuição deles no território. Além disso, o artigo comporta a visualização de áreas com potencial para a ocupação deles através de uma perspectiva da estatística, o que contribui para a aproximação destas disciplinas fornecendo subsídios para os avanços nas pesquisas arqueológicas que tem como ênfase o estado de São Paulo.

AGRADECIMENTOS

Aos Professores Dr. Luís Gustavo Esteves, Dr. Antônio Carlos Pedroso de Lima, Dr. Carlos Alberto de Bragança Pereira, Dr. Victor Fossaluza, aos alunos de graduação Virgínia Carrara e Adriano Gonçalves Lima do Departamento de Estatística do Instituto de Matemática e Estatística da Universidade de São Paulo, do Centro de Consultoria em Estatística Aplicada – o CEA, que foram os responsáveis por apresentar o programa R e auxiliar na sua aplicação à arqueologia. O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) com bolsa de Doutorado para GCP entre 2014 e 2018. Ao CNPq, Conselho Nacional de Desenvolvimento Científico e Tecnológico – Brasil, pela bolsa de produtividade em pesquisa de MO (processo n° 302163/2017-4) e MCA (processo n° 310373/2016-6).

REFERÊNCIAS BIBLIOGRÁFICAS

- AFONSO, Marisa Coutinho 2005. *Um olhar para a arqueologia pré-histórica do Estado de São Paulo.* 2005. Tese de Livre Docência. Museu de Arqueologia e Etnologia, Universidade de São Paulo, São Paulo.
- ARAUJO, Astolfo Gomes de Mello. 2007. A tradição cerâmica Itararé-Taquara: características, área de ocorrência e algumas hipóteses sobre a expansão dos grupos Jê no sudeste do Brasil. *Revista de Arqueologia.* 20: 09-38.
- ASSIS, Valéria Soares de. 1996. *Da espacialidade Tupinambá*. Dissertação de Mestrado Pontifícia Universidade Católica do Rio Grande do Sul.
- BAXTER, Mike; COOL, Hilary. 2016. Basic statistical graphics for archaeology with R: life beyond Excel.
- BINFORD, Lewis. 1962. A new method of calculating dates from Kaolin pipe stem samples. *Southeastern Archaeological Conference*, Newsletter 9(1): 19-21.
- BONOMO, Mariano; ANGRIZANI, Rodrigo Costa; APOLINAIRE, Eduardo; NOELLI, Francisco Silva. 2015 A model for the Guaraní expansion in the La Plata Basin and littoral zone of southern Brazil. *Quaternary International*. 356, 54-73
- BROCHADO, José Proenza. 1973. Migraciones que difundieron la tradición alfarera Tupiguarani. *Relaciones 7*, Sociedad Argentina de Antropologia, Buenos Aires. p. 7-39.
- BROCHADO, José Proenza. 1984. An ecological modelo f the preado f pottery and agriculture into Easter South America. Tese de Doutorado, University of Illinois at Urbana, Champaign, Ann Arbor UMI.
- BUCK, Caitlin E., KENWORTHY, J.B., LITTON, C.D. and SMITH, A.F., 1991. Combining archaeological and radiocarbon information: a Bayesian approach to calibration. *Antiquity*, 65 (249), pp.808-821.
- CARLSON, David. 2017. Quantitative methods in archaeology using R. Cambridge University Press, Texas A & M University.
- CLARKE, David. 1968. Analytical archaeology. Londres, Methuen.
- CORREA, Ângelo. 2014. *Pindorama de Mboîa e Îakaré:* continuidade e mudança na trajetória das populações Tupi. Tese de Doutorado. MAE/USP.
- COWGILL, George L., 1970. Some sampling and reliability problems in archaeology. In: *Archeologie et calculateurs: Problemes semiologiques et mathematiques* (pp. 161-175).
- DUNNELL, Robert. 1980. Evolutionary theory and archaeology. In *Advances in archaeological method and theory* (pp. 35-99). Academic Press.
- DUNNELL, Robert; DANCEY, William, 1983. The siteless survey: a regional scale data collection strategy. In *Advances in archaeological method and theory* (pp. 267-287). Academic Press.
- ESTEVES, Luís Gustavo; FOSSALUZA, Victor; LIMA, Adriano Gonçalves; CARRARA, Virgínia T. 2017. Relatório de Análise Estatística sobre o Projeto: "Arqueologia Paulista e o marcador cerâmico como delimitador de fronteira étnica: um estudo da região sul do estado de São Paulo", Relatório final. (RAE CEA 17P23).
- GASPAR, Maria Dulce. 1991. Aspectos da organização social de um grupo de pescadores, coletores e caçadores: região compreendida entre Ilha Grande e o delta do Paraíba do Sul, Estado do Rio de Janeiro. Tese de doutorado. MAE/USP.
- GASPAR, Maria Dulce. 1996. Análise da bibliografia sobre pescadores, coletores e caçadores que ocuparam o Estado do Rio de Janeiro. *Revista do Museu de Arqueologia e Etnologia*. SP. 6: 337-367.

- LANDA, Beatriz dos Santos. 1995. *A mulher Guarani: atividades e cultura material.* Dissertação de Mestrado Rio Grande do Sul, Pontifícia Universidade Católica do Rio Grande do Sul.
- LATHRAP, Donald Ward. 1970. The Upper Amazon, New York, Praeger.
- MAGESTE, Leandro Elias Canaan. 2017. Cronologia e variabilidade: os ceramistas Tupiguarani da Zona da Mata Mineira e Complexo Lagunar de Araruama. Tese de Doutorado São Paulo.
- MARWICK, Ben. 2016. Computational reproducibility in archaeological research: basic principles and case study of their Implementation. *Journal of Archaeological Method and Theory*. New York, January.
- MEDRI, Waldir. 2011. *Análise exploratória de dados*. Centro de Ciências Exatas CCE. Departamento de Estatística. Londrina. Disponível em: http://www.uel.br/pos/estatisticaeducacao/textos_didaticos/especializacao_estatistica.pdf Acessado em 18/01/2017.
- MEGGERS, Betty Jane; EVANS, Clifford. 1973. A reconstituição da pré-história amazônica: algumas considerações teóricas. IN: O Museu Goeldi no ano do Sesquicentenário, Publicações avulsas, Belém, 20, p. 51-69.
- MONTICELLI, Gislene. 1995. Análises das informações obtidas com os Mbyá-Guarani sobre suas antigas vasilhas de cerâmica. *Revista CEPA*. Santa Cruz do Sul, Universidade Federal do Rio Grande do Sul, 23(29): 233-239.
- MOTA, Lucio Tadeu. 1998. *O aço, a cruz e a terra: índios e brancos no Paraná provincial (1853-1889).*Tese de Doutorado Faculdade de Ciências e Letras, Universidade Estadual de São Paulo, Assis, 531f.
- MOTA, Lúcio Tadeu. 2016. A passagem e a presença dos Jê Meridionais por São Paulo e Paraná: uma reflexão etno-histórica. *Revista do Museu de Arqueologia e Etnologia*, n. 27, p. 135-157.
- MOTA, Lúcio Tadeu. 2000. Os índios Kaingang e seus territórios nos campos do Brasil meridional na metade do século passado. In: Lúcio Tadeu Mota; Francisco Silva Noelli; Kimiye Tommasino. (Org.). *Uri e Wāxi: estudos interdisciplinares dos Kaingang*. 1ªed.Londrina: EDUEL, v. 1, p. 81-190.
- MOTA, Lucio Tadeu; ASSIS, Valéria Soares de. 2008. Populações indígenas no Brasil: histórias, culturas e relações interculturais. Maringá, Eduem.
- MOTA, Lucio Tadeu; NOELLI, Francisco Silva. 1999. Exploração e guerra de conquista dos territórios indígenas nos vales dos rios Tibagi, Ivaí e Piquiri. IN: *DIAS, R. B.; GONÇALVES; J. H. R.* (*Org.*) *Maringá e o Norte do Paraná*. Maringá, Eduem, p. 21-50.
- MORAES, Camila Azevedo de. 2007. Arqueologia tupi no Nordeste de São Paulo: um estudo de variabilidade artefatual. Dissertação de Mestrado. MAE/USP. São Paulo.
- MYERS, Oliver. 1950. Some Applications of Statistics to Archaeology. Published by the Service des Antiquités de l'Egypte, Cairo, 1950. In *Proceedings of the Prehistoric Society* (Vol. 16, pp. 200-201). Cambridge University Press.
- NOELLI, Francisco Silva. 1993. Sem Tekoá não há Tekó Em busca de um modelo etnoarqueológico da aldeia e da subsistência Guarani e sua aplicação a uma área de domínio no delta do rio Jacuí, Rio Grande do Sul. Dissertação de Mestrado. Porto Alegre, Pontifícia Universidade Católica do Rio Grande do Sul,
- OKUMURA, Mercedes; ARAUJO, Astolfo Gomes de Mello. 2019. Archaeology, biology, and borrowing: A critical examination of Geometric Morphometrics in Archaeology. *Journal of Archaeological Science*, 101, pp.149-158.
- OTÁROLA-CASTILLO, Erick.; TORQUATO, Melissa. 2018. Bayesian statistics in archaeology. *Annual Review of Anthropology*, 47, pp.435-453.

- PEARSON, Jessica; GROVE, Matthew. 2013. Counting sheep: sample size and statistical inference in stable isotope analysis and palaeodietary reconstruction. *World Archaeology*, *45*(3), pp. 373-387.
- PEREZ, Glauco Constantino. 2018. Arqueologia Paulista e o marcador cerâmico como delimitador de fronteira étnica: um estudo das regiões sul e oeste do Estado de São Paulo. Tese de doutorado. MAE/USP. Disponível em: http://www.teses.usp.br/teses/disponiveis/71/71131/tde-05042018-092642/pt-br.php
- PEREZ, Glauco Constantino; AFONSO, Marisa Coutinho; MOTA, Lúcio Tadeu. 2018^a. Métodos de análise espacial para sítios arqueológicos: um modelo preditivo para o Estado de São Paulo. *Cadernos do LepaArq.* V. XV. N.30. jul-dec.. P.98-120.
- PINHEIRO, Niminon Suzel. 1999. Vanuíre. Conquista, colonização e indegenismo: oeste paulista, 1912-1967. Tese de Doutorado. Universidade Estadual Paulista. UNESP/SP.
- REDMAN, Charles. 1974. Archeological sampling strategies. Reading, Massachusetts: Addison-Wesley.
- RODRIGUES, Robson Antônio. 2007. Os caçadores-ceramistas do sertão paulista: um estudo etnoarqueológico da ocupação Kaingang no vale do Rio Feio/Aguaí. 2007. Tese de Doutorado. Museu de Arqueologia e Etnologia, Universidade de São Paulo. São Paulo.
- SAIA, Soraya Elaine Marques Gouveia. 2006. Reconstrução paleoambiental (vegetação e clima) no Quaternário tardio com base em estudo multi/interdisciplinar no Vale do Ribeira (sul do estado de São Paulo). Mestrado. USP, São Paulo.
- SCHMITZ, Pedro Ignácio. 2007. Índios guaranis, Kaingang e Xokleng: territórios indígenas e fronteiras. IN: MOREIRA, L. F. V. (Org.). *Instituições, fronteiras e políticas na História Sulamericana*. Curitiba: Editora Juruá. p. 129-148.
- SOARES, André Luís. 1997. Guarani: organização social e arqueologia. *Coleção Arqueologia 4*, EDIPUCRS, RS,
- SPAULDING, Albert. 1953. Statistical techniques for the discovery of artifact types. *American antiquity*, 18(4), pp.305-313.
- SUSNIK, Branislava. 1975. Dispersón Tupi-Guarani prehistorica. Ensayo analítico. Museo Etnográfico Andrés Barbero Ltda., Asunción, Paraguay.