

Centro de Estatística Aplicada

Relatório de Análise Estatística

RAE-CEA-20P07

RELATÓRIO DE ANÁLISE ESTATÍSTICA SOBRE O PROJETO:

“Um estudo sobre a distribuição dos sítios arqueológicos da Tradição Aratu na paisagem capixaba”

Alexandre Galvão Patriota

Danilo Vieira Silva

Rodrigo Passos Martins

São Paulo, 30 de junho de 2020

CENTRO DE ESTATÍSTICA APLICADA - CEA – USP

TÍTULO: Relatório de Análise Estatística sobre o Projeto: “Um estudo sobre a distribuição dos sítios arqueológicos da Tradição Aratu na paisagem capixaba”.

PESQUISADORA: Dionne Miranda Azevedo Erler

ORIENTADORA: Profa. Dra. Marisa Coutinho Afonso

INSTITUIÇÃO: Museu de Arqueologia e Etnologia (MAE-USP)

FINALIDADE DO PROJETO: Mestrado

RESPONSÁVEIS PELA ANÁLISE: Alexandre Galvão Patriota

Danilo Vieira Silva

Rodrigo Passos Martins

REFERÊNCIA DESTE TRABALHO: PATRIOTA, A.G.; SILVA, D.V. e MARTINS, R.P.

Relatório de análise estatística sobre o projeto: “Um estudo sobre a distribuição dos sítios arqueológicos da Tradição Aratu na paisagem capixaba”. São Paulo, IME-USP, 2020. (RAE–CEA-20P07).

REFERÊNCIAS BIBLIOGRÁFICAS:

BUSSAB, W. O.; MORETTIN, P. A. (2013). **Estatística Básica**. 8.ed. Editora Saraiva, 540p.

ESTEVES, L. G., FOSSALUZA, V., LIMA, A. G.; CARRARA, V. T. **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”**, São Paulo, IME-USP, 2020. (RAE–CEA-17P23).

FUNARI, Pedro Paulo Abreu (2003). **Arqueologia**. 3.ed. São Paulo: Contexto. 126p.

PAULINO, C.D.; SINGER, J.M. (2006). **Análise de dados categorizados**. São Paulo: Blücher.

PEROTA, C. (1975). As datações de C14 dos sítios do Espírito Santo. **Revista de Cultura da UFES**, 4, 15-16.

PEROTA, C. (1971). Dados parciais sobre a arqueologia norte espírito-santense **Programa Nacional de Pesquisa Arqueológicas: Resultados preliminares do quarto ano. Publicações Avulsas do Museu Paraense Emílio Goeldi**, 149-158.

PEROTA, C. (1974). Resultados preliminares sobre a arqueologia da região central do estado do Espírito Santo. **Programa Nacional de Pesquisa Arqueológicas: Resultados preliminares do quinto ano. Publicações Avulsas do Museu Paraense Emílio Goeldi**, 127-139.

POLETO, F. Z. **Funções em S-Plus/R para análises de MLGs**. Disponível em: <<http://www.poleto.com/funcoes.html>> Acesso em: 01 de junho de 2020.

SEN, P.K., SINGER, J.M.; PEDROSO DE LIMA, A.C. (2010). **From Finite Sample to Asymptotic Methods in Statistics**. New York, Cambridge University Press.

VENABLES, W. N. and RIPLEY, B. D. **Modern Applied Statistics with S. Fourth edition**. Springer, 2002.

PROGRAMAS COMPUTACIONAIS UTILIZADOS:

Google Docs (versão 2020);

Google Sheets (versão 2020);

R para Windows (versão 3.6.3).

TÉCNICAS ESTATÍSTICAS UTILIZADAS:

Análise Descritiva Unidimensional (03:010).

ÁREA DE APLICAÇÃO:

Ciências Físicas e Geoestatística (14:010).

Sumário

1. Introdução	7
2. Objetivos	8
3. Descrição do estudo	8
4. Descrição das variáveis	9
5. Análise descritiva	11
6. Análise inferencial.....	24
7. Conclusões	41
APÊNDICE A	44
APÊNDICE B	53
APÊNDICE C	58
APÊNDICE D	68
ANEXO.....	78

1. Introdução

A Arqueologia estuda os modos de vida das sociedades humanas, tanto do passado como do presente, a partir da análise de objetos materiais (FUNARI, 2003). Quando o interesse está em estudar determinado aspecto de filiações culturais do passado, os objetos analisados são materiais culturais encontrados em sítios arqueológicos, onde ficam preservados esses materiais.

Estudamos a localização de filiações culturais do estado do Espírito Santo e suas preferências a partir da análise da distribuição espacial de sítios arqueológicos no estado considerando variáveis geoespaciais e culturais. As filiações culturais estudadas são: Sambaqui, tradição Itaipu, tradição Aratu, tradição Tupiguarani e tradição Una. Essas tradições são identificadas por meio do modelo da divisão da Pré-história proposto por Perota (1971; 1974; 1975) para o Espírito Santo. O modelo da Pré-história do Espírito Santo é dividido em dois períodos: o pré-cerâmico (7000 a.C. até 1050 a.C.) relacionado com as filiações Sambaqui e caçadores-coletores (tradição Itaipu) do litoral, e o cerâmico (desde 1050 a.C.) relacionado com as filiações ceramistas e de contato: tradição Aratu, tradição Tupiguarani e tradição Una.

O principal indicativo na identificação da filiação cultural de um sítio arqueológico é o tipo de material cultural que o sítio apresenta, porém, não é determinante. Outros aspectos, como decoração do material, idade do material e localização do sítio, também devem ser levados em consideração na hora da identificação, esse fato é o que motiva a análise estatística desse problema.

Na Seção 2, apresentamos os objetivos principais deste estudo, bem como os estudos de caso posteriores. Na Seção 3, temos a descrição do estudo e como foram obtidos os dados. Na Seção 4, fazemos a descrição das variáveis utilizadas no estudo mostrando para cada variável a respectiva unidade de medida ou categoria, além de abreviações. Na Seção 5, realizamos a análise descritiva das variáveis descritas na seção anterior e, na Seção 6, realizamos a análise inferencial. Na Seção 7, apontamos as conclusões das análises realizadas. Por fim, no Apêndice A, apresentamos algumas figuras da distribuição espacial dos sítios arqueológicos no Espírito Santo, no Apêndice B, apresentamos algumas tabelas descritivas, no Apêndice C, apresentamos a

metodologia utilizada na análise inferencial, no Apêndice D, apresentamos a análise de resíduos para os modelos e simulações e, no Anexo, apresentamos alguns exemplos de materiais culturais associados a cada filiação.

2. Objetivos

Analizar a distribuição espacial dos sítios arqueológicos no estado do Espírito Santo considerando variáveis geoespaciais e culturais, visando estabelecer possíveis preferências das filiações para escolha dos locais de seus assentamentos. A partir disso:

1. Identificar os sítios arqueológicos de filiação não definida em uma das filiações observadas;
2. Fazer o estudo de caso acerca da relação da ocupação do espaço pela tradição Aratu com marcos paisagísticos da região lagunar do município de Linhares, o qual se localiza no litoral norte do Espírito Santo, onde se encontra a maior concentração de sítios da tradição Aratu.

3. Descrição do estudo

Foi feito um levantamento de dados bibliográficos e cartográficos visando rastrear e examinar características dos sítios arqueológicos do Espírito Santo. Os dados bibliográficos foram analisados minuciosamente e por meio deles foram obtidas as informações sobre as variáveis culturais dos sítios, depois foram processados no Sistema de Informações Geográficas (SIG) para a criação de um banco de dados georreferenciado. Foram considerados 490 sítios arqueológicos, sendo 251 sítios com filiação definida e 239 sítios com filiação não definida.

Para confecção dos produtos cartográficos foi utilizado o SIG com o software ArcGIS 10.6, adotando o sistema de coordenadas Universal Transversa de Mercator (UTM), a projeção cartográfica o Datum Sirgas 2000 e zona 24S. Essas informações foram convertidas posteriormente para o respectivo sistema de coordenadas com Latitude e Longitude dando origem às variáveis geoespaciais dos sítios.

4. Descrição das variáveis

Variáveis sobre a coordenada de referência:

- **Sítio:** nome único de cada um dos 490 sítios arqueológicos;
- **Coordenada E (metros):** distância em relação ao meridiano Central;
- **Coordenada N (metros):** distância em relação à Linha do Equador;
- **Latitude (grau, minuto e segundo):** distância à Linha do Equador medida ao longo do meridiano de Greenwich;
- **Longitude (grau, minuto e segundo):** é a distância ao meridiano de Greenwich medida ao longo da Linha do Equador.

Variáveis sobre a descrição do material cultural:

- **Período:** Pré-colonial e Pré-colonial/Contato da ocupação humana;
- **Componente:** Unicomponencial e Multicomponencial na composição do sítio;
- **Filiação cultural:** Sambaqui, Itaipu, Aratu, Tupiguarani, Una, Não Definido, Sambaqui/Aratu, Sambaqui/Tupiguarani, Tupiguarani/Aratu, Itaipu/Tupiguarani e Itaipu/Aratu;
- **Material cultural:** CP, Lítico, Malacológico, CP/Lítico, CP/CC, CP/Lítico/Malacológico/CC, Lítico/Malacológico/Ósseo, Lítico/CC, Todos, CC/Malacológico/Ósseo/Lítico e CC/Lítico/Ósseo/Malacológico;
 - CC: Cerâmica colonial;
 - CP: Cerâmica pré-colonial.
- **Tipo cerâmico pré-colonial:** Simples, Dec.Pintada, Dec.Plástica, Simples/Dec.Plástica, Simples/Dec.Pintada, Todos, Não especificado e Não aplicável.
 - Dec.: Decoração.

Variáveis sobre características geoespaciais:

- **Área (m²):** área do sítio demarcada pelo arqueólogo;
- **Geomorfologia:** Chapadas/Planaltos/Patamares, Depressão, Escarpas/Reversos, Planícies (Marinhas/Fluviomarinhas/Fluviolacustres), Tabuleiros costeiros e Planícies (Fluviais/Fluviolacustres);
- **Distância (quilômetros):** distância até a principal fonte de recurso hídrico;
- **Clima:** Oceânico, Oceânico temperado, Tropical savana, Tropical úmido, Tropical monçônico e Não especificado;
- **Elevação (metros):** Elevação < 30, Elevação > 30;
- **Unidade:** APA (Costa das Algas), APA (Conceição da Barra), APA (Sepetiba), APA (Maciço Central), APA (Guanandy), Não especificado, ARIE (Degredo), PE (Itaúnas), PE (Paulo César Vinha), PNM (Jacarenema), RE (Manguezais-Piraquê-Açu/Piraquê-Mirim), RE (Ilha do Lameiro) e RVS (Santa Cruz);
 - APA: Área de Proteção Ambiental;
 - ARIE: Área de Relevante Interesse Ecológico;
 - PE: Parque Estadual;
 - PNM: Parque Nacional Municipal;
 - RE: Reserva Ecológica;
 - RVS: Refúgio de Vida Silvestre.
- **Uso:** Água, Alagado, Cultura, Floresta, Manguezal, Não especificado, Pastagem, Praia, Restinga, Silvicultura e Urbano;
- **Vegetação:** Área formações pioneiras, Flor. estacional semidecidual, Flor. ombrófila densa e Água;
 - Flor.: Floresta;
- **Geologia:** Holoceno, Inferior, Água, Paleoceno, Pleistoceno, Médio e Não aplicável;
- **Pedologia:** Latossolos, Argilossolos, Neossolos, Chernossolos, Gleissolos, Espodossolos, Cambissolo, Solos Halomórficos e Não informado.

5. Análise descritiva

A unidade amostral é o **Sítio** arqueológico e a variável resposta é a **Filiação cultural**. Apresentamos a análise descritiva dos dados para as variáveis numéricas e categóricas (BUSSAB e MORETTIN, 2013).

Tabela 1. Distribuição da **Filiação cultural**

	n	%
Sambaqui	97	19,8%
Itaipu	14	2,9%
Aratu	29	5,9%
Tupiguarani	91	18,6%
Una	14	2,9%
Itaipu/Tupiguarani	1	0,2%
Sambaqui/Tupiguarani	3	0,6%
Tupiguarani/Aratu	2	0,4%
Não Definido	239	48,8%
Total	490	100%

A Tabela 1 mostra a quantidade e a proporção de sítios arqueológicos de cada **Filiação cultural**. Notamos a grande quantidade de sítios com **Filiação cultural** igual a “Não Definido”.

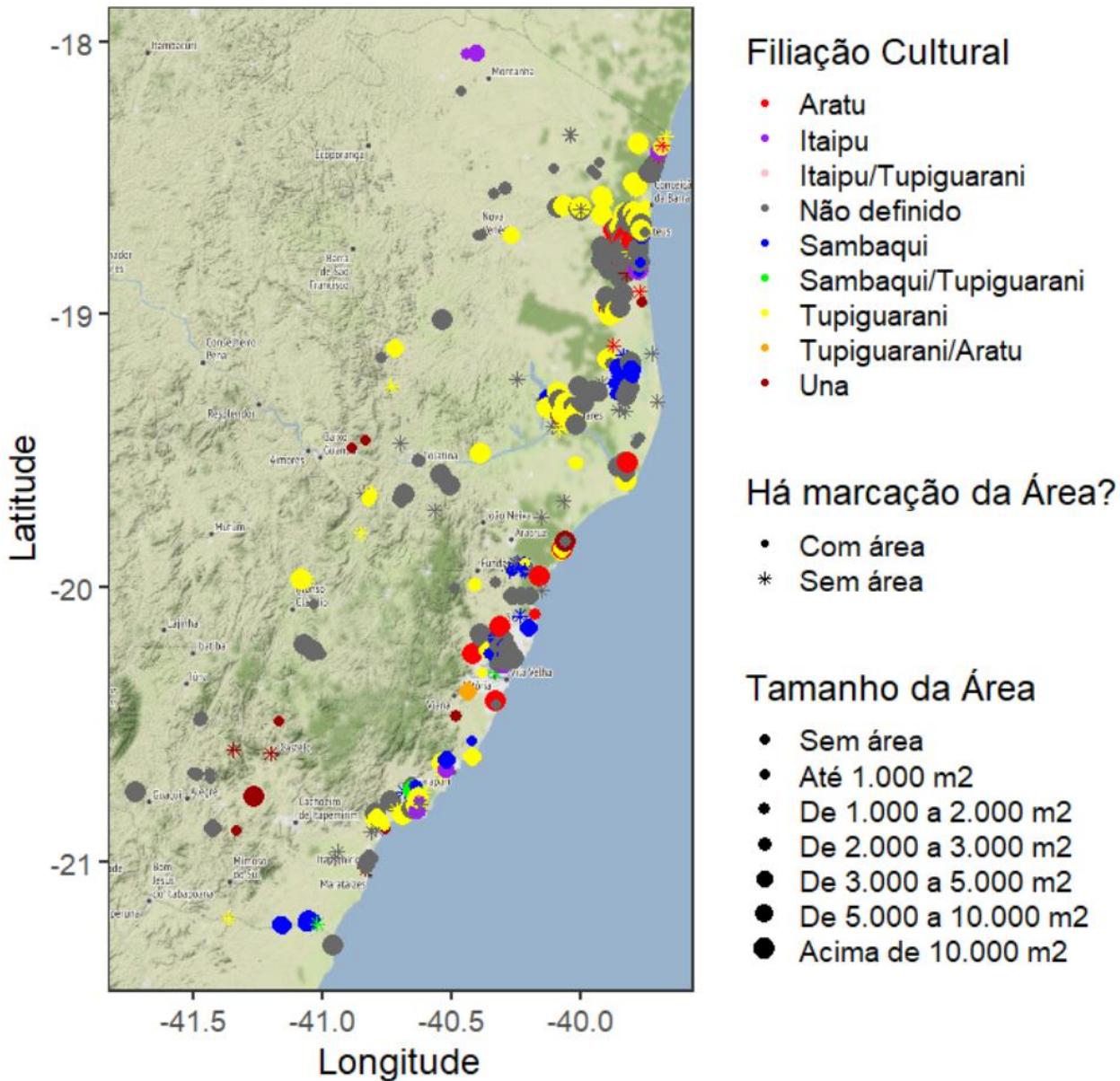


Figura 1. Distribuição da **Filiação cultural** de acordo com a localização

A Figura 1 mostra os sítios distribuídos espacialmente no Espírito Santo em que o tamanho da **Área** de cada sítio está indicado de acordo com a tamanho do círculo que o representa. Indicamos os sítios sem **Área** especificada com uma estrela. Para essa representação, as variáveis **Coordenada E** e **Coordenada N** foram transformadas em **Latitude** e **Longitude**. Notamos que a região Lagunar de Linhares (Longitude: -40° E

e Latitude: -19° N) apresenta muitos sítios Aratu e Tupiguarani, essa região faz parte do nosso estudo de caso (Seção 6.3). Notamos também que a filiação Una apresenta preferência por regiões mais ao interior do Estado. As Figuras A.1 a A.8 apresentam essa mesma representação considerando apenas os sítios de cada **Filiação cultural** e os sítios Não Definido.

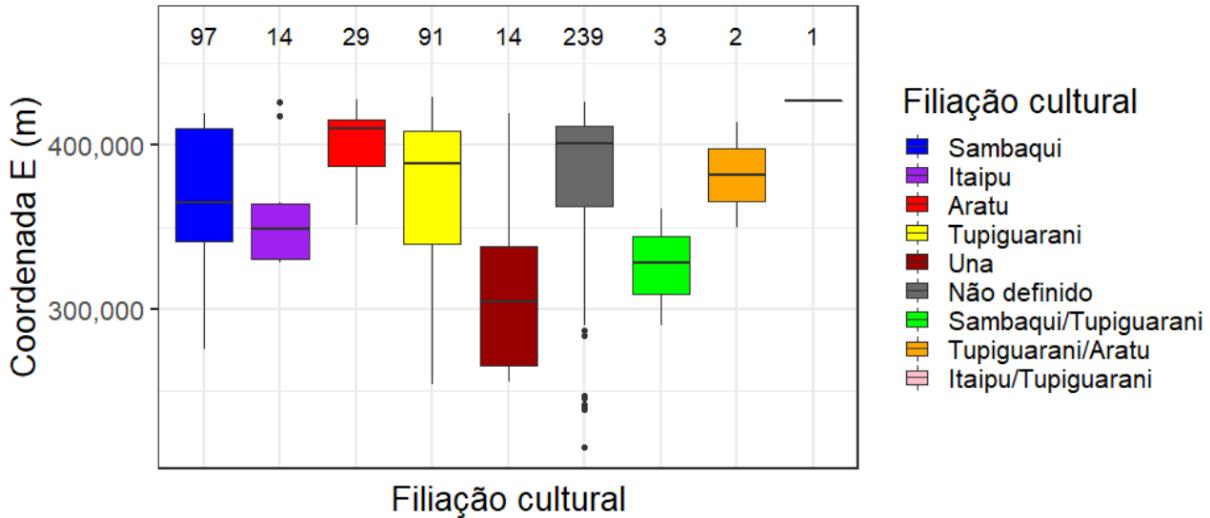


Figura 2. Box plot da **Coordenada E**

A Figura 2 mostra que as filiações culturais Una e Sambaqui/Tupiguarani apresentam menores valores de **Coordenada E** em relação às demais. Esse fato mostra que essas tradições são mais interioranas do que as outras no território capixaba (Aratu e Tupiguarani já são mais litorâneas). Temos três *box plots* com menos de 10 sítios e alguns *outliers* nas filiações Itaipu e Não Definido.

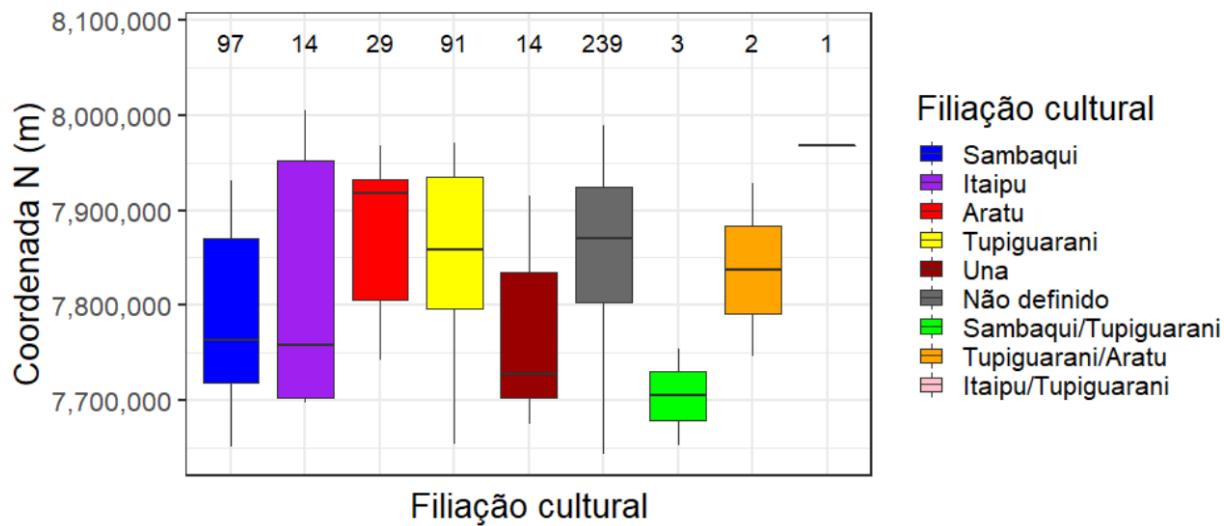


Figura 3. Box plot da **Coordenada N**

Já a Figura 3 apresenta as ocupações de Filiação cultural Itaipu mais distribuídas entre o norte e o sul do estado, enquanto as outras tradições apresentam uma dispersão mais controlada, sendo Sambaqui e Una mais espalhadas do que Aratu e Tupiguarani. Temos três *box plots* com menos de 10 sítios.

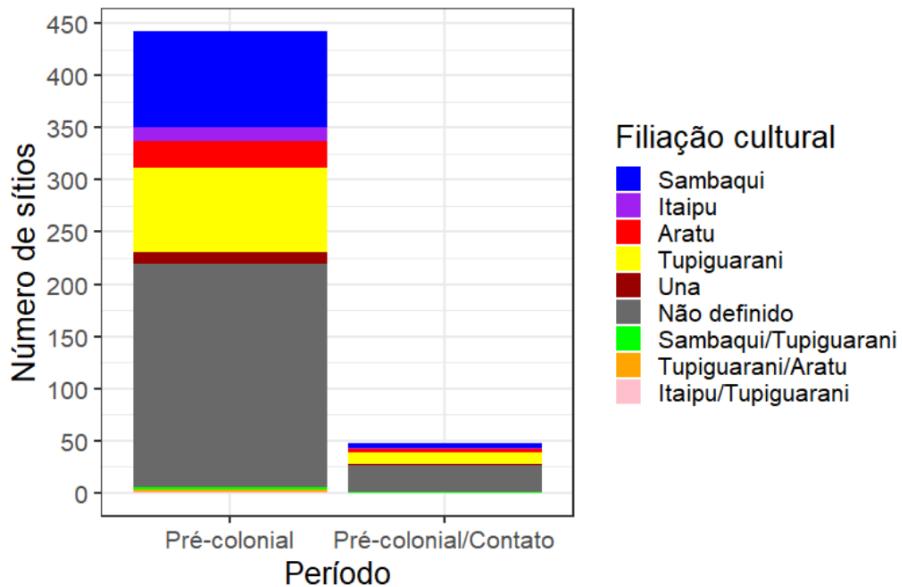


Figura 4. Gráfico de barras de **Período**

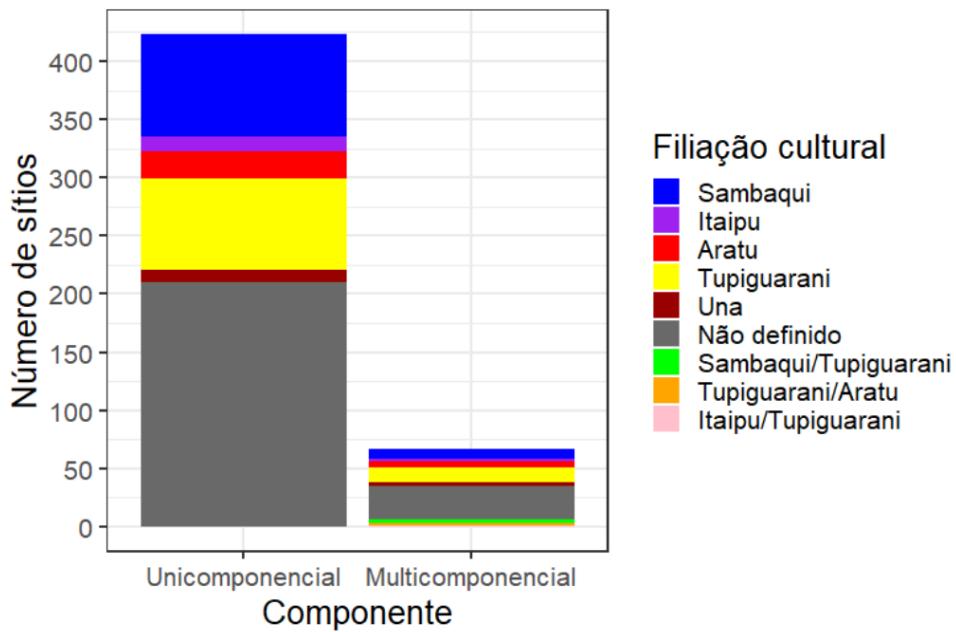


Figura 5. Gráfico de barras de **Componente**

A Figura 4 mostra que a maioria dos sítios apresentam período Pré-colonial e a distribuição dos sítios neste período é muito semelhante à dos sítios no período Pré-colonial/Contato. Não há prevalência exclusiva de nenhum **Período** para alguma **Filiação cultural**.

A Figura 5 mostra que a maioria dos sítios apresentam componente Unicomponencial e a distribuição dos sítios nesta componente é muito semelhante à dos sítios na componente Multicomponencial. Notamos que não existem sítios com componente Unicomponencial classificados com mais de uma **Filiação cultural**, mas existem sítios com componente Multicomponencial com apenas uma **Filiação cultural**. Isso ocorre porque os sítios Multicomponencial apresentaram **Material cultural** do **Período** Colonial/histórico, mas os sítios com o **Período** Colonial/histórico não compõem o cerne da análise.

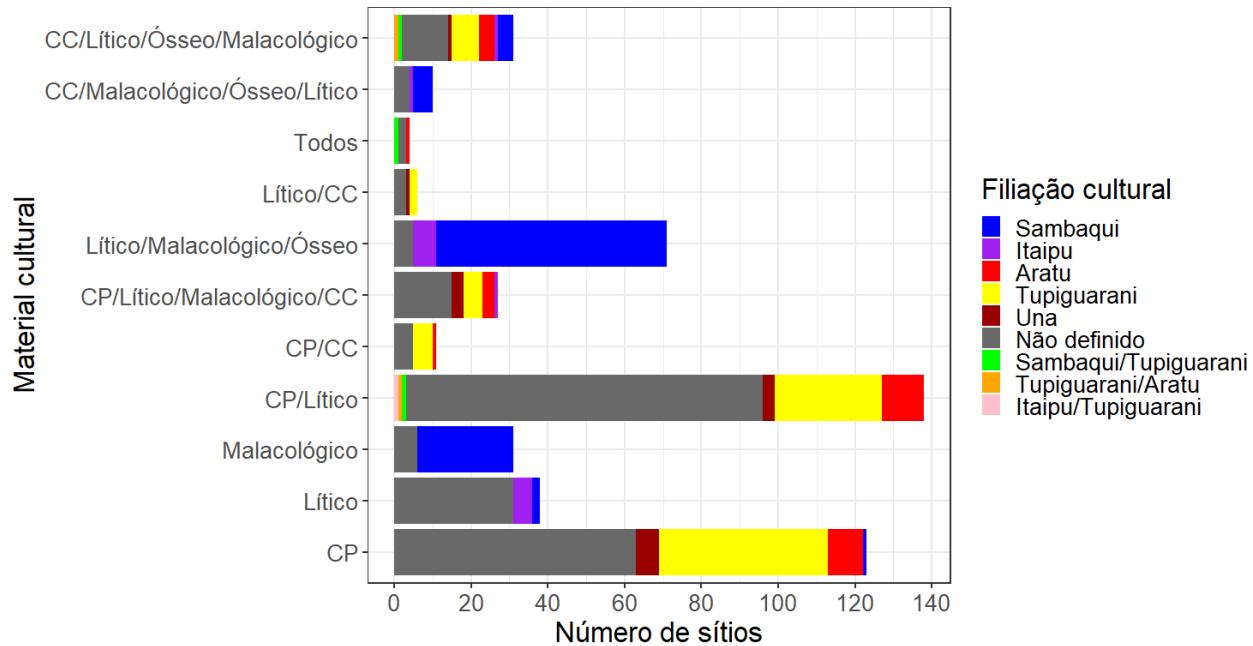


Figura 6. Gráfico de barras de Material cultural

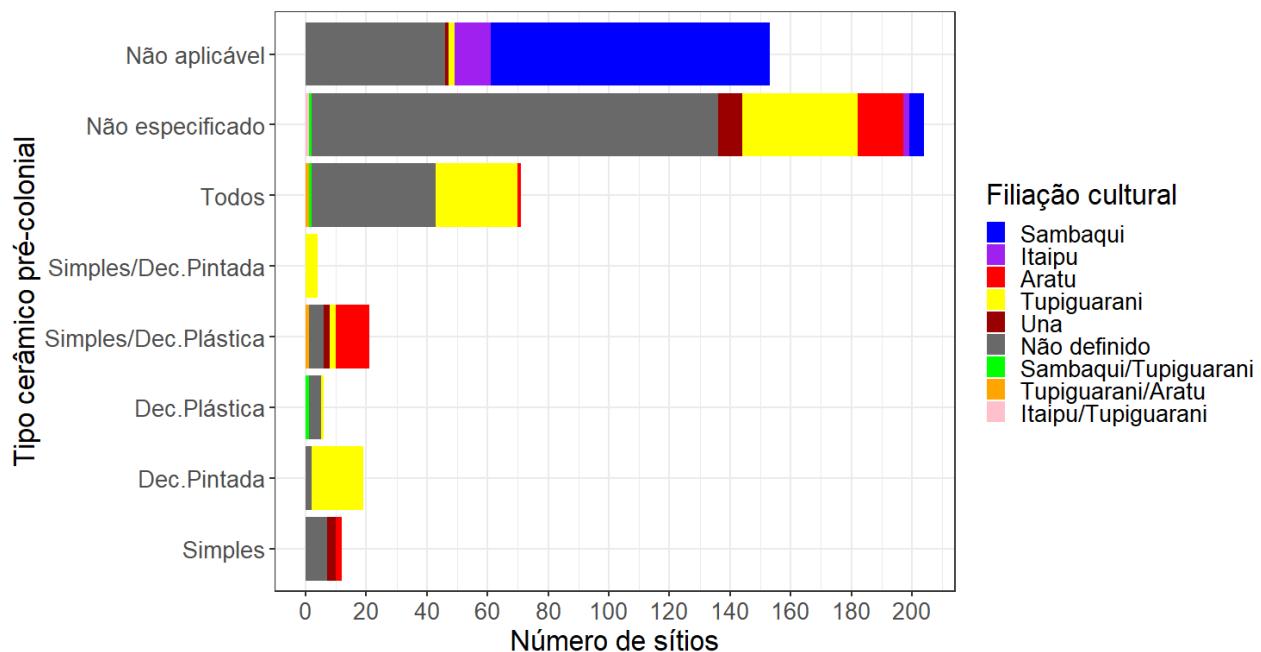


Figura 7. Gráfico de barras de Tipo cerâmico pré-colonial

Na Figura 6, os sítios das filiações não ceramistas, Sambaqui e Itaipu, possuem **Material cultural** Malacológico ou Lítico/Malacológico/Ósseo, enquanto que os sítios das filiações ceramistas possuem, majoritariamente, CP ou CP/Lítico. As Figuras AN.1 à AN. 14 apresentam alguns exemplos desses materiais.

A Figura 7 destaca o fato de que os sítios das filiações não ceramistas, Sambaqui e Itaipu, não estão aptos para uma análise do **Tipo cerâmico pré-colonial**, uma vez que, não possuem esse tipo de material em sua **Filiação cultural**. Já entre os sítios elegíveis nesse tipo de análise, os compostos por grupos ceramistas, notamos a preferência de Dec.Pintada pela filiação Tupiguarani e a preferência de Dec.Plástica pela filiação Aratu.

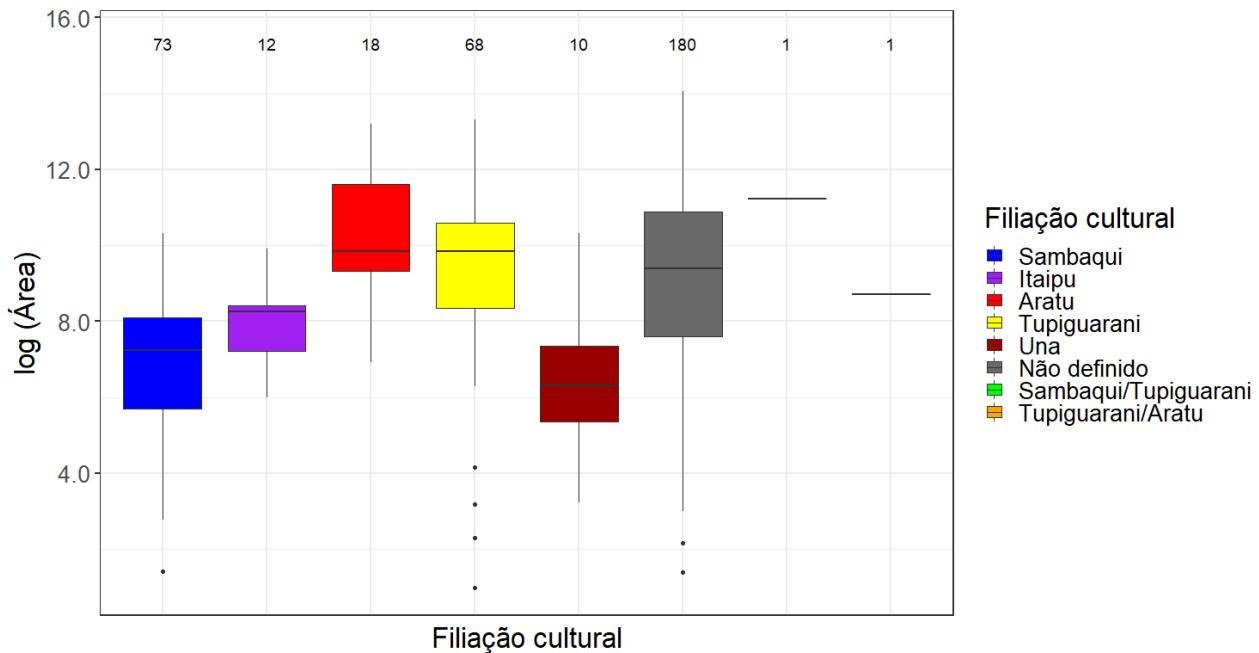


Figura 8. Box plot da Área

Para a montagem da Figura 8, retiramos 127 sítios com **Área** igual a zero e calculamos o log desta informação para os restantes. Pela imagem, percebemos que os maiores sítios são das filiações Aratu e Tupiguarani, sendo que 75% deles detêm uma **Área** maior que em 75% dos sítios das outras tradições (1º quartil é igual ou superior ao

3º quartil). Temos dois *box plots* com menos de 10 sítios e alguns *outliers* nas filiações Sambaqui, Tupiguarani e Não Definido.

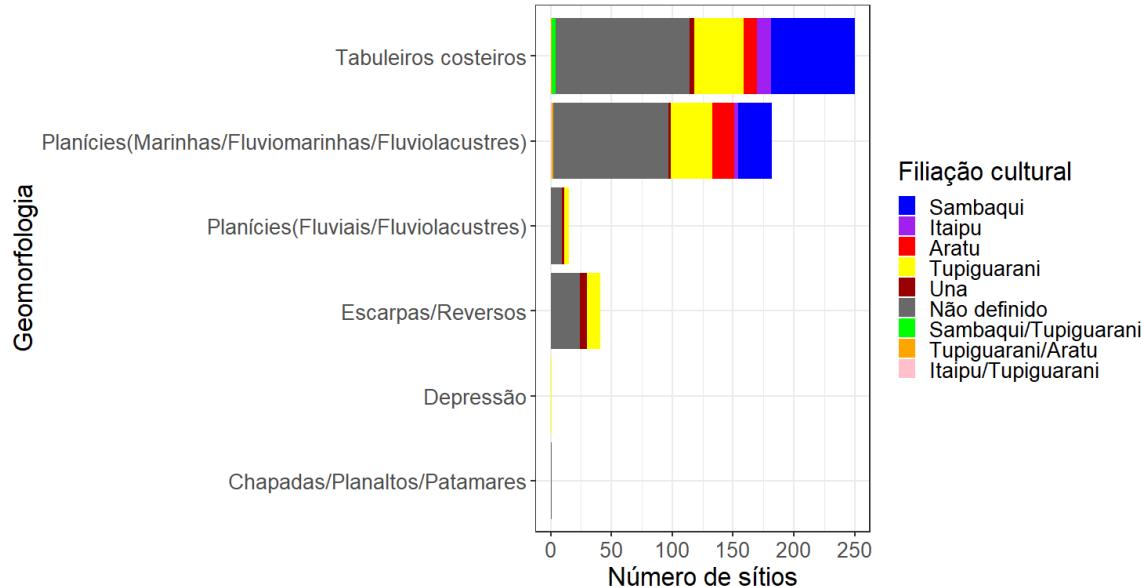


Figura 9. Gráfico de barras de **Geomorfologia**

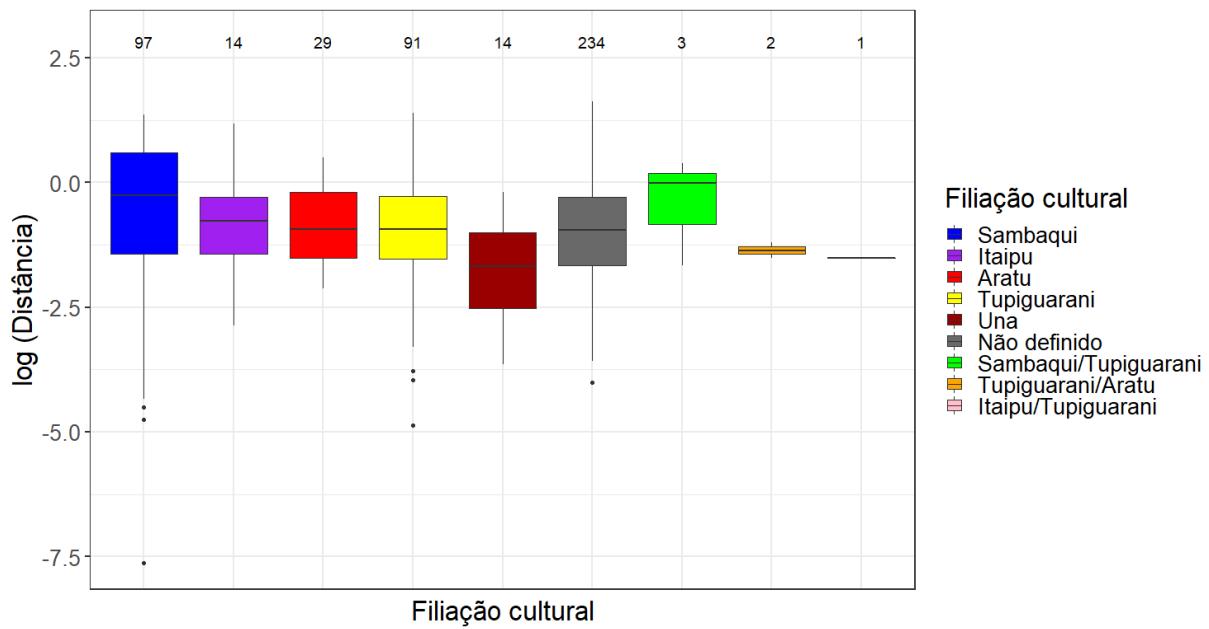


Figura 10. Box plot da **Distância**

A Figura 9 apresenta o gráfico de barras para os níveis de **Geomorfologia**. Esses níveis correspondem a diversas classificações para a superfície terrestre. Por conta da uniformidade da superfície do Espírito Santo e da distribuição diversa dos sítios ao longo do estado, essa informação não ajuda a discriminar as filiações.

Para o desenvolvimento da Figura 10 removemos 5 sítios com **Distância** zero e aplicamos a função logarítmica nas distâncias dos demais sítios. Com isso, temos que as medidas são parecidas entre as diversas tradições, exceto para a tradição Una que apresenta as menores distâncias. Porém, é importante notar que apenas as filiações Sambaqui e Sambaqui/Tupiguarani possuem logaritmo da **Distância** maior que zero, ou seja, **Distância** maior que 1 quilômetro do recurso hídrico mais próximo. Temos três *box plots* com menos de 10 sítios e alguns *outliers* nas filiações Sambaqui, Tupiguarani e Não Definido.

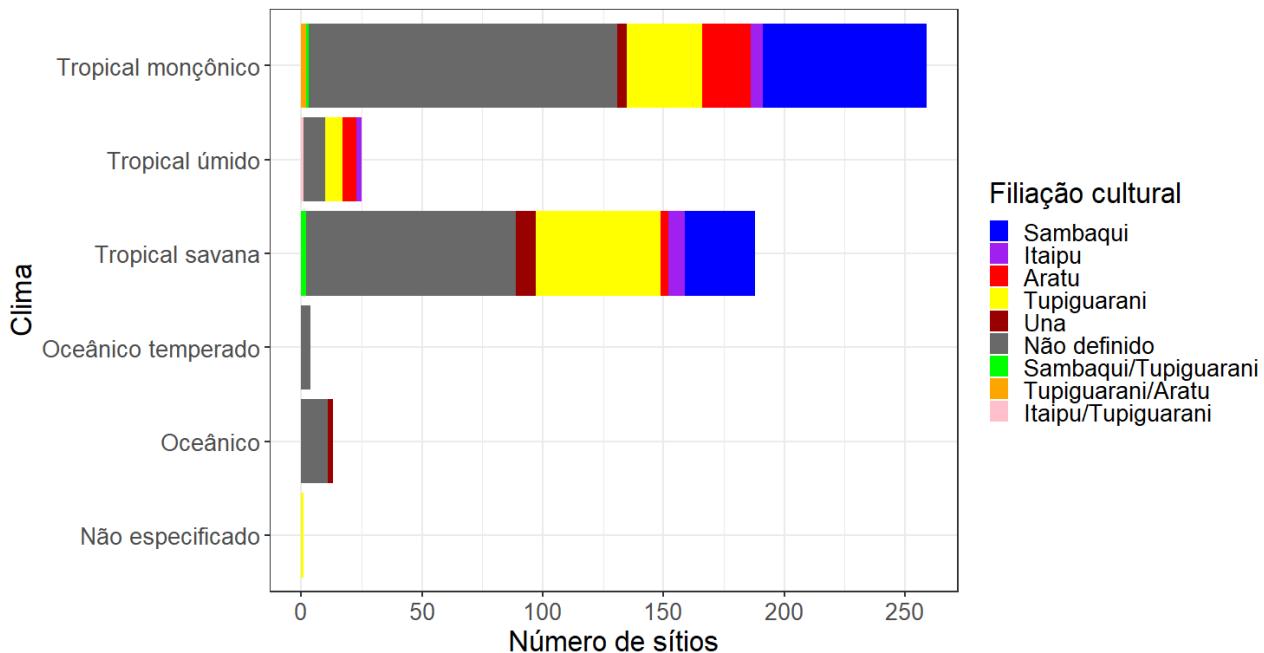


Figura 11. Gráfico de barras de **Clima**

A Figura 11 mostra que a maioria dos sítios possuem clima Tropical monçônico ou savana, não revelando preferência de nenhuma **Filiação cultural** por **Clima**.

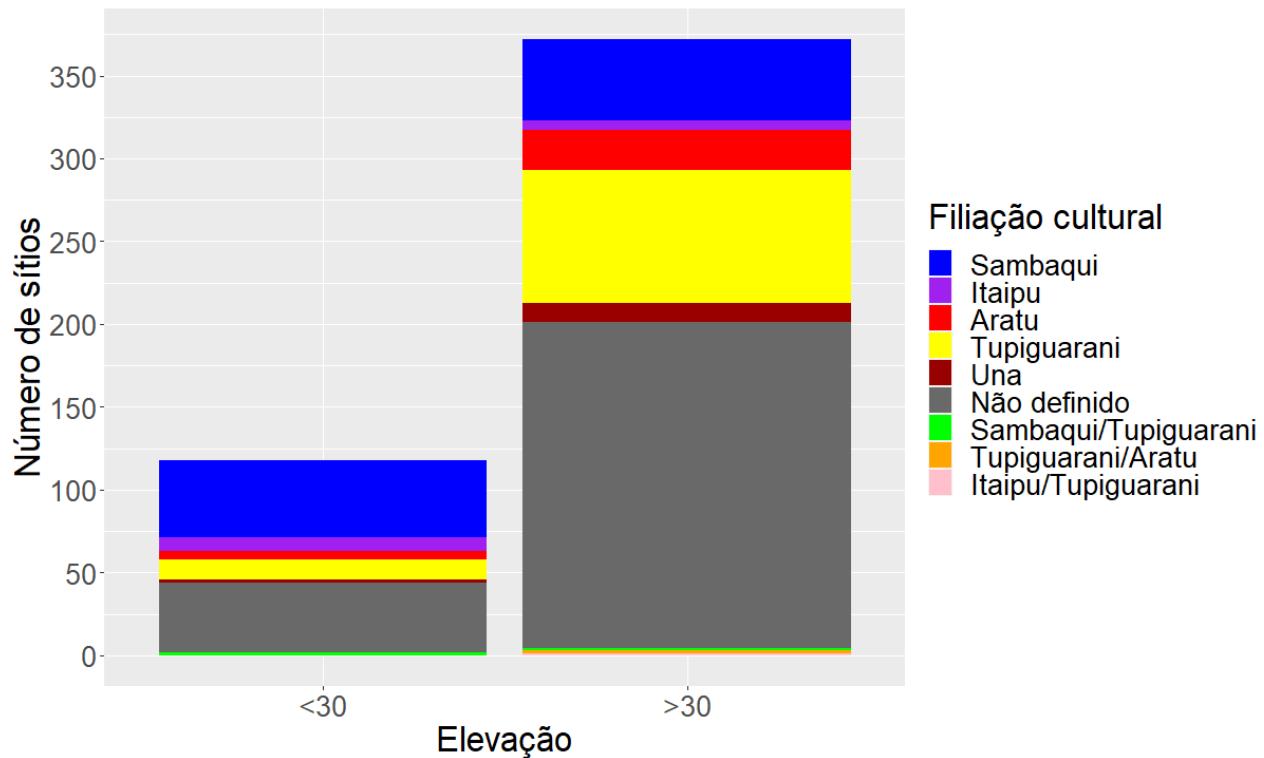


Figura 12. Gráfico de barras de **Elevação**

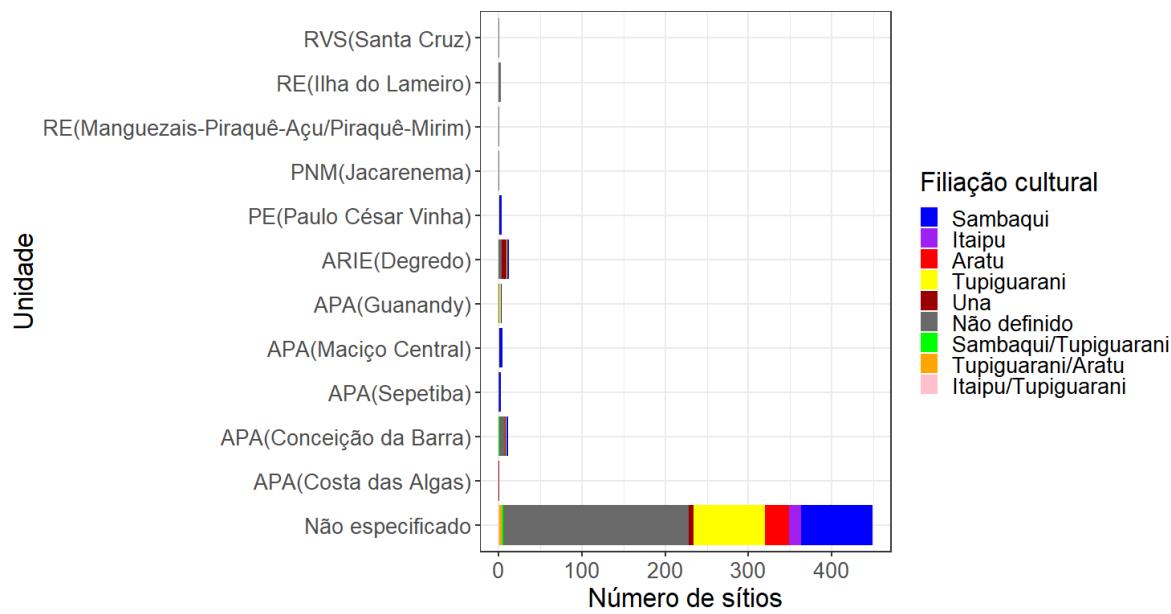


Figura 13. Gráfico de barras de **Unidade**

A informação da Figura 12 ratifica o que pode ser concluído nos mapas da Figura 1 e das Figuras A.1 a A.8, ou seja, uma ocupação litorânea representativa (**Elevação baixa**). Percebemos também que a maioria dos sítios da filiação Una possuem **Elevação maior** que 30 metros.

A Figura 13 apresenta a falta de especificação da **Unidade** na maioria dos sítios analisados, assim essa informação não ajuda a discriminar as filiações.

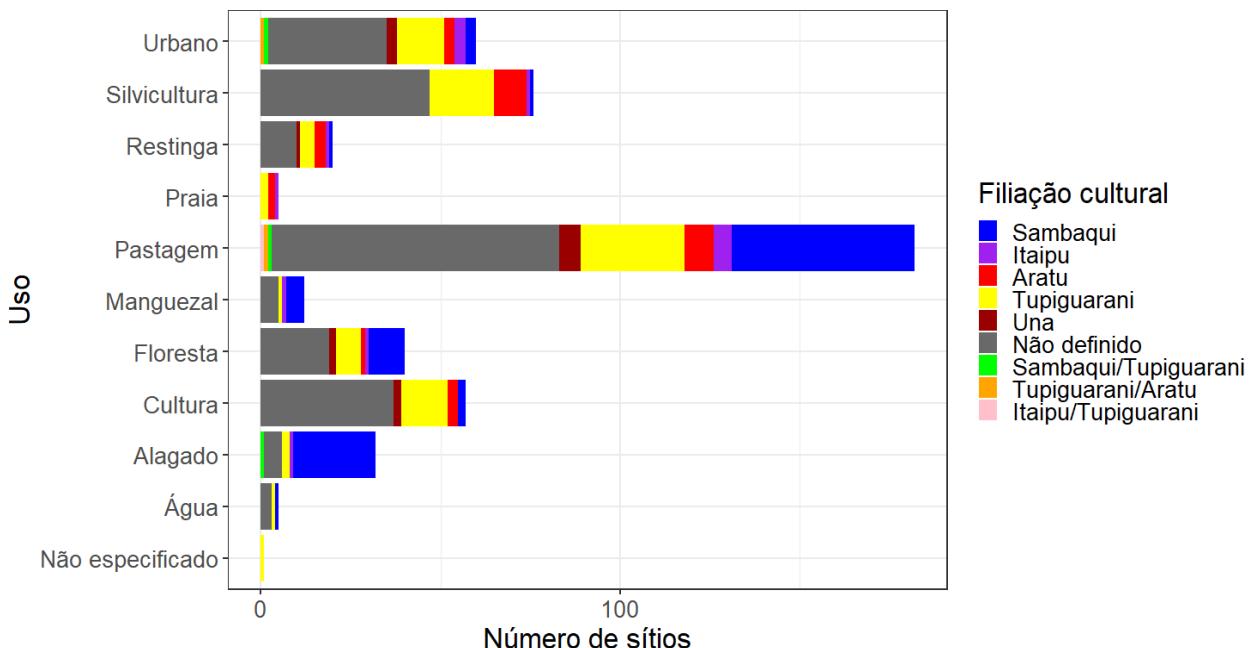


Figura 14. Gráfico de barras de **Uso**

A Figura 14 mostra que não há relação clara entre o **Uso** e a **Filiação cultural** e que a maioria dos sítios são usados atualmente para Pastagem. Notamos também que os sítios Não Definido aparecem em todos os níveis de **Uso** exceto em Praia e Não especificado.

A Figura 15 não mostra uma relação clara entre a **Filiação cultural** e a **Vegetação**, muito por conta da homogeneidade da paisagem capixaba. Mas é possível notar a preferência da filiação Itaipu pela **Vegetação** de Flor. ombrófila densa.

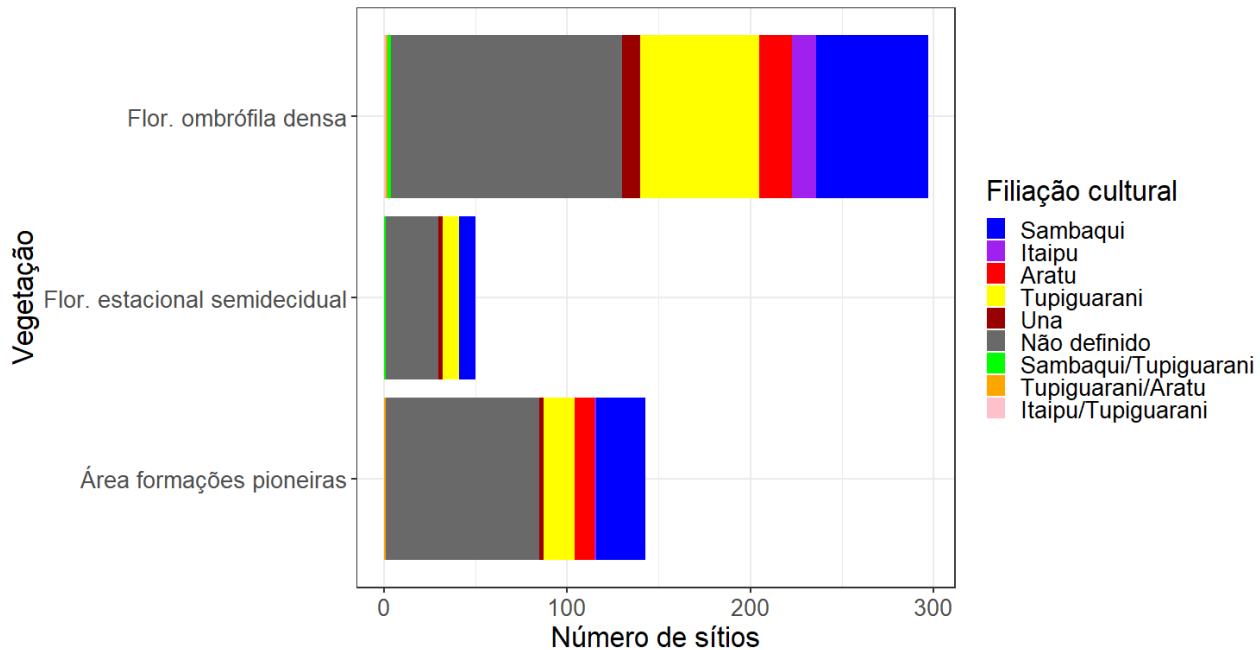


Figura 15. Gráfico de barras de **Vegetação**

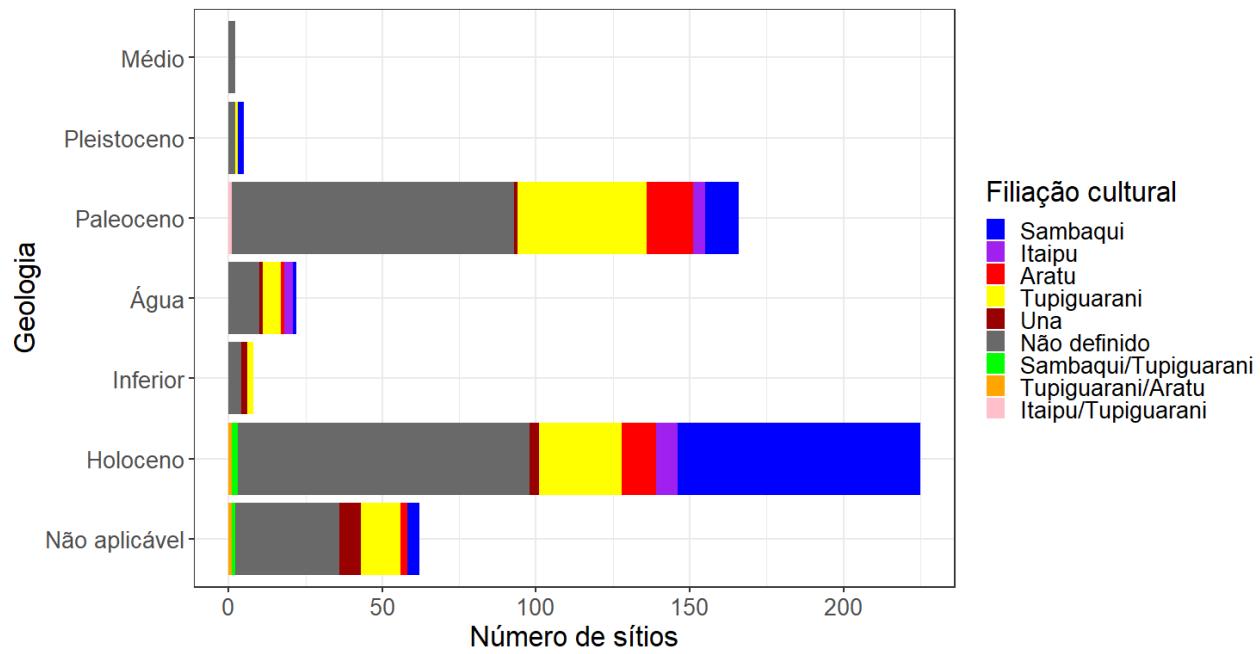


Figura 16. Gráfico de barras de **Geologia**

A Figura 16 apresenta o gráfico de barras para os níveis de **Geologia**. Esses níveis correspondem a diversos períodos de ocupações da crosta terrestre. Observamos uma ocupação predominante no Holoceno (crosta terrestre mais antiga) e no Paleoceno (crosta terrestre mais recente). Notamos também que os sítios Não Definido aparecem em todos os níveis de **Geologia**.

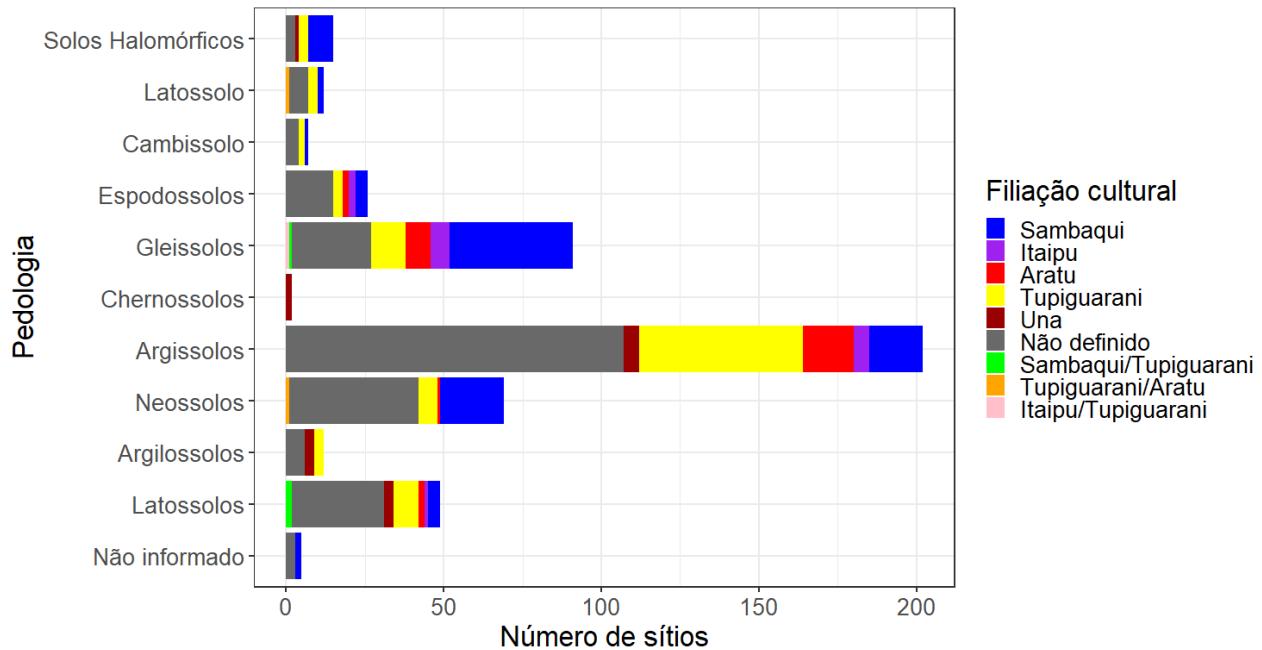


Figura 17. Gráfico de barras de **Pedologia**

A Figura 17 mostra que não há preferência clara de alguma **Pedologia** por alguma **Filiação cultural** e que a maioria dos sítios são Argissolos. Notamos também que os sítios Não Definido aparecem em todos os níveis de **Pedologia** exceto em Chernossolos.

6. Análise inferencial

O objetivo deste projeto é classificar os sítios arqueológicos de filiação Não Definido (ver Seção 4) em uma das filiações observadas e fazer o estudo de caso acerca da ocupação de Linhares pela tradição Aratu. Como observado na Análise descritiva (Seção 5), não há observações em todos os níveis de todas as variáveis considerando as oito filiações culturais. Essa situação afeta o uso do conjunto de dados para a modelagem no R, pois os dados são transformados em matrizes numéricas e esse excesso de zeros prejudica os cálculos matriciais. Por conta disso, não consideramos em nossa análise classificações com mais de uma filiação cultural (ex: Sambaqui/Tupiguarani). Assim, excluímos 6 sítios do conjunto de dados e classificamos os sítios arqueológicos de filiação Não Definido em apenas uma das cinco filiações culturais: Aratu, Tupiguarani, Una, Itaipu e Sambaqui.

Como também observamos na Análise descritiva, algumas variáveis apresentam muitos níveis, então, a fim de reduzir o número de níveis das variáveis para termos observações em todos os níveis de todas as variáveis, com o auxílio da pesquisadora, excluímos e padronizamos algumas variáveis e agrupamos os níveis de outras. Para mais detalhes desse processo, consulte o Apêndice B.

Apesar das modificações supramencionadas, os níveis de algumas variáveis ainda não apresentavam observações, mas restaram somente zeros estruturais. Essa característica configura um problema computacional, pois o R trabalha com matrizes numéricas e o excesso de zeros acarreta em imprecisões nas inversões de matrizes, no cálculo de determinante, entre outros. No caso, mesmo aumentando o número de sítios arqueológicos, esses níveis dessas variáveis ainda não apresentariam observações, mantendo zeros em matrizes. Para resolver esse problema, utilizamos dois modelos para classificação (Paulino & Singer, 2006): um binomial para os sítios com filiações não ceramistas (Itaipu e Sambaqui) e um modelo trinomial para os sítios com filiações ceramistas (Aratu, Tupiguarani e Una). Após a aplicação do procedimento, os níveis de todas as variáveis têm pelo menos uma observação considerando as cinco filiações culturais. Para mais detalhes da metodologia proposta, veja o Apêndice C.

6.1 Modelo binomial

Para os sítios não ceramistas (Itaipu e Sambaqui) consideramos 110 sítios e aplicamos um modelo multinomial com duas categorias (binomial) utilizando a parametrização de logito de referência com a filiação Itaipu na referência. As variáveis explicativas testadas inicialmente nesse modelo foram: **Distância, Elevação, Coordenada E, Coordenada N, Pedologia, Geomorfologia, Período e Material cultural**. Consideramos como referência os sítios com as características apresentadas na Tabela 2 a qual descreve para cada variável o seu respectivo nível ou valor de referência. Essa referência está relacionada ao intercepto do modelo e sua interpretação está nesta subseção.

Tabela 2. Valores e níveis para o sítio de referência para os Não Cerâmicos (unidades amostrais do modelo binomial)

Variável	Valor/Nível de referência
Distância padronizada	0
Elevação	< 30
Coordenada E padronizada	0
Coordenada N padronizada	0
Pedologia	Outro
Geomorfologia	Tabuleiros/Depressão/Escarpas/Reversos
Período	Pré-colonial
Material cultural	Outro

Ajustamos um modelo considerando todas as variáveis e, a partir dele, encontramos o modelo final pelo método *stepwise* que utiliza o Critério de Informação de Akaike (*Akaike Information Criterion – AIC*). Para mais detalhes sobre a estimação do modelo final, veja o Apêndice C.

Os parâmetros estimados para o modelo final estão na Tabela 3, juntamente com os respectivos erros-padrão e os valores-p para a hipótese de que o valor do parâmetro é zero. Também foi realizado o teste conjunto em que se postula que os parâmetros são iguais a zero exceto o intercepto e, ao nível de significância 5%, rejeitamos a hipótese nula (com valor-p <0,001) de que o modelo final poderia ser reduzido somente ao intercepto. Fizemos simulações para verificar se o tamanho da amostra era

suficientemente grande para se utilizar a teoria assintótica nos testes de hipóteses. Todas as simulações tiveram resultados satisfatórios mostrando que o tamanho da amostra é adequado. Para mais detalhes sobre as simulações, veja o Apêndice D.

Tabela 3. Estimativas dos parâmetros estimados para o modelo binomial com erros-padrão associados e valor-p do respectivo teste de hipótese individual

Variável	Estimativa (Erro-padrão)	Valor-p
Intercepto	0,42 (0,35)	0,479
Distância padronizada	0,86 (0,73)	0,060*
Elevação > 30	1,29 (0,67)	0,103
Coordenada E padronizada	2,30 (0,21)	0,004**
Coordenada N padronizada	-2,60 (0,63)	0,002**

Cada parâmetro pode ser interpretado individualmente em função da chance ou da razão de chances de o sítio ser classificado como Itaipu em relação a ser classificado como Sambaqui. Essa interpretação é usada para entender a influência de cada variável na classificação do sítio. A interpretação de cada coeficiente estimado está descrita abaixo.

- **Intercepto**

A chance de ser Sambaqui é $\exp(0,42) = 1,52$ para um sítio cujas características se encontram na Tabela 2 (níveis de referência), ou seja, a probabilidade de ser Sambaqui é 52% (pois, $1 - \exp(0,42) = 0,52$) maior que a probabilidade de ser Itaipu para um sítio nos níveis de referência.

- **Distância padronizada**

A chance do sítio ser Sambaqui com acréscimo em 1 unidade na Distância padronizada é 136% (i.e., $1 - \exp(0,86) = 1,36$) maior que a chance do sítio ser Sambaqui sem esse acréscimo na Distância padronizada, mantendo-se todas as outras variáveis constantes.

- **Elevação > 30**

A chance do sítio ser Sambaqui com Elevação maior que 30m é 263% (i.e., $1 - \exp(1,29) = 2,63$) maior que a chance do sítio ser Sambaqui com Elevação igual a referência, mantendo-se todas as outras variáveis constantes.

- **Coordenada E padronizada**

A chance do sítio ser Sambaqui com acréscimo em 1 unidade na Coordenada E padronizada é 897% (i.e., $1 - \exp(2,30) = 8,97$) maior que a chance do sítio ser Sambaqui sem esse acréscimo na Coordenada E padronizada, mantendo-se todas as outras variáveis constantes.

- **Coordenada N padronizada**

A chance do sítio ser Sambaqui com acréscimo em 1 unidade na Coordenada N padronizada é 93% (i.e., $1 - \exp(-2,60) = -0,93$) menor que a chance do sítio ser Sambaqui sem esse acréscimo na Coordenada N padronizada, mantendo-se todas as outras variáveis constantes.

Estudamos os resíduos dos modelos para verificar o ajuste. Os resultados se mostraram satisfatórios, indicando que o modelo está bem ajustado. Para mais detalhes, veja o Apêndice D.

Verificamos a qualidade do modelo observando o número de acertos na matriz de confusão (Tabela 4). Para tal, definimos a regra de classificação, com o auxílio da pesquisadora, olhando a probabilidade de o sítio pertencer a cada filiação. O Quadro 1 apresenta a regra de classificação sugerida aos sítios Não Cerâmicos. Os erros de

classificações entre as filiações têm o mesmo peso. Sendo assim, maximizamos a diagonal principal da matriz contida na Tabela 4, apresentamos os resultados para a pesquisadora e em conjunto desenvolvemos as regras contidas no Quadro 1.

Quadro 1. Método de classificação sugerido dos sítios Não Cerâmicos, segundo o modelo binomial

Filiação Cultural	Regra de classificação
Itaipu	Probabilidade de o sítio ser Itaipu, segundo o modelo binomial, maior que 20%.
Sambaqui	Probabilidade de o sítio ser Sambaqui, segundo o modelo binomial, maior que 80%.

Tabela 4. Matriz de confusão para o modelo final para não ceramistas

		PREDITO	
REAL		Itaipu	Sambaqui
Itaipu	Itaipu	5	9
	Sambaqui	14	82

Segundo a regra do Quadro 1, fizemos um mapa com a classificação dos sítios sem filiação definida anteriormente. Utilizamos apenas os sítios Não Definido cuja classificação é não ceramista. A Figura 18 apresenta o mapa em questão.

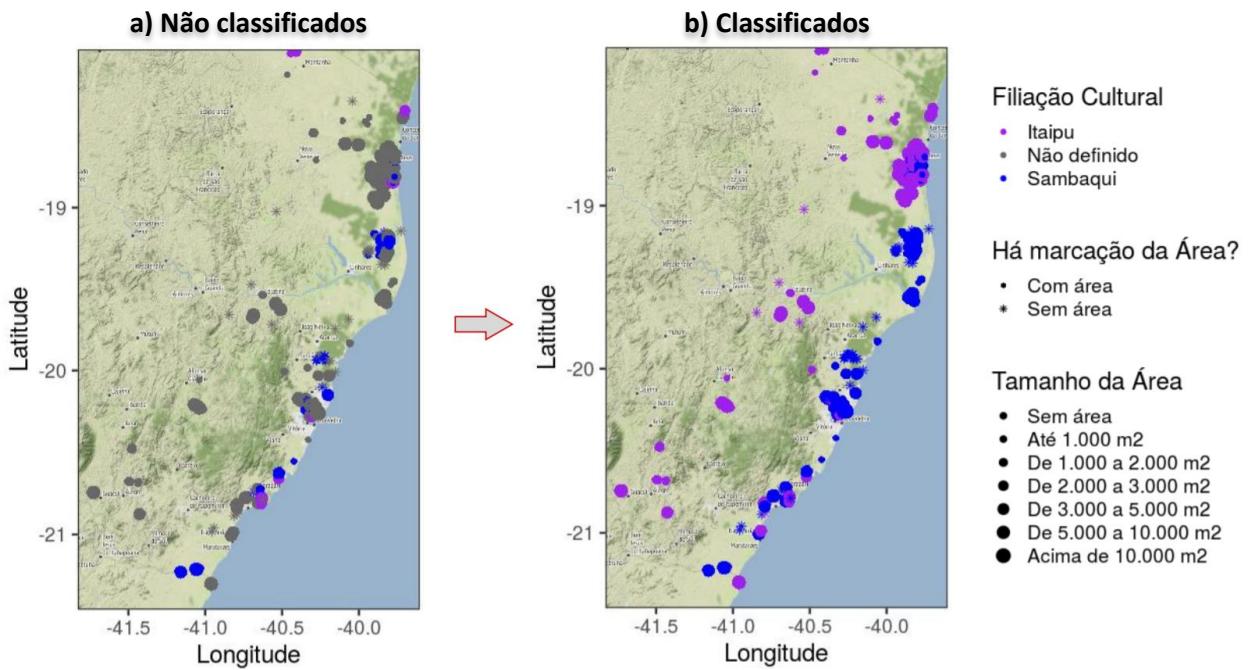


Figura 18. Classificação dos sítios Não Definido por meio das probabilidades do modelo binomial

6.2 Modelo trinomial

Para os sítios com filiações ceramistas (Aratu, Tupiguarani e Una) consideramos 135 sítios e aplicamos um modelo multinomial com três categorias (trinomial) utilizando a parametrização de logito de referência com a filiação Tupiguarani na referência. As variáveis utilizadas nesse modelo foram: **Distância, Elevação, Coordenada E, Coordenada N, Pedologia, Geomorfologia, Período, Material cultural, Componente** e **Tipo cerâmico**. Consideramos como referência os sítios com as características apresentadas na Tabela 5 a qual descreve para cada variável o seu respectivo nível ou valor de referência. Essa referência está relacionada ao intercepto do modelo e sua interpretação está nessa subseção.

Tabela 5. Valores e níveis para o sítio de referência para os Cerâmicos (unidades amostrais do modelo trinomial)

Variável	Valor/Nível de referência
Distância padronizada	0
Elevação	<30
Coordenada E padronizada	0
Coordenada N padronizada	0
Pedologia	Outro
Geomorfologia	Tabuleiros/Depressão/Escarpas/Reversos
Período	Pré-colonial
Material cultural	Outro
Componente	Unicomponencial
Tipo cerâmico	Simples/Dec.Pintada

Ajustamos um modelo considerando todas as variáveis e, a partir dele, encontramos o modelo final pelo método *stepwise* que utiliza o AIC. Para mais detalhes sobre a obtenção do modelo final, veja o Apêndice C.

Os parâmetros estimados para o modelo final estão na Tabela 6, juntamente com os respectivos erros-padrão e os valores-p para a hipótese de que o parâmetro é zero. Também foi realizado o teste conjunto em que se postula que todos os parâmetros são iguais a zero exceto o intercepto, ao nível de significância 5%, rejeitamos a hipótese nula (com valor-p <0,001) de que o modelo final poderia ser reduzido somente ao intercepto. Como anteriormente, verificamos também por meio de histogramas das simulações, o

tamanho amostral para ver se esse é adequado para o uso da teoria assintótica nos testes de hipóteses. O resultado foi bom para a maioria apesar de alguns indicarem possíveis problemas de convergência (e.g. Figura D.9).

Tabela 6. Estimativas dos parâmetros estimados para o modelo binomial com erros-padrão associados e valor-p do respectivo teste de hipótese individual

Variável	Filiação associada	Estimativa (Erro-padrão)	Valor-p
Intercepto	Aratu	-3,19 (0,82)	<0,001***
Distância padronizada	Aratu	-0,51 (0,18)	0,228
Coordenada E padronizada	Aratu	3,13 (1,60)	0,013**
Coordenada N padronizada	Aratu	-1,02 (0,60)	0,188
Geomorfologia-Planície	Aratu	-0,62 (0,76)	0,477
Tipo cerâmico pré-colonial-Simples/Dec.Plástica	Aratu	3,27 (1,38)	0,005**
Tipo cerâmico pré-colonial-Todos/Não aplicável	Aratu	-2,33 (1,79)	0,081*
Tipo cerâmico pré-colonial-Não especificado	Aratu	0,79 (0,75)	0,357
Material cultural-Ceramica/Outro	Aratu	1,63 (0,37)	<0,001***
Intercepto	Una	-6,13 (2,32)	<0,001***
Distância padronizada	Una	-1,59 (0,77)	0,070*
Coordenada E padronizada	Una	-1,57 (0,35)	0,008**
Coordenada N padronizada	Una	-1,34 (0,70)	0,108
Geomorfologia-Planícies	Una	4,14 (2,81)	0,013**
Tipo cerâmico pré-colonial-Simples/Dec.Plástica	Una	5,09 (2,69)	0,002**
Tipo cerâmico pré-colonial-Todos/Não aplicável	Una	-0,87 (2,35)	0,566
Tipo cerâmico pré-colonial-Não especificado	Una	0,50 (1,03)	0,621
Material cultural-Ceramica/Outro	Una	-0,08 (0,76)	0,923

As interpretações dos parâmetros estimados para o modelo trinomial são análogas aos parâmetros estimados para o modelo binomial:

- **Intercepto (Aratu)**

A chance de ser Aratu em relação a ser Tupiguarani é $\exp(-3,19) = 0,04$ para um sítio com todas as variáveis nos valores e níveis de referência, ou seja, a probabilidade de ser Aratu é 96% (pois $1 - \exp(-3,19) = -0,96$) menor que a probabilidade de ser Tupiguarani para esse sítio.

- **Distância padronizada (Aratu)**

A chance do sítio ser Aratu em relação a ser Tupiguarani com acréscimo em 1 unidade na Distância padronizada é 40% (i.e., $1 - \exp(-0,51) = -0,40$) menor que a chance do sítio ser Aratu em relação a ser Tupiguarani sem esse acréscimo na Distância padronizada mantendo-se todas as outras variáveis constantes.

- **Coordenada E padronizada (Aratu)**

A chance do sítio ser Aratu em relação a ser Tupiguarani com acréscimo em 1 unidade na Coordenada E padronizada é 2187% (i.e., $1 - \exp(3,13) = 21,87$) maior que a chance do sítio ser Aratu em relação a ser Tupiguarani sem esse acréscimo na Coordenada E padronizada mantendo-se todas as outras variáveis constantes.

- **Coordenada N padronizada (Aratu)**

A chance do sítio ser Aratu em relação a ser Tupiguarani com acréscimo em 1 unidade na Coordenada N padronizada é 64% (i.e., $1 - \exp(-1,02) = -0,64$) menor que a chance do sítio ser Aratu em relação a ser Tupiguarani sem esse acréscimo na Coordenada N padronizada mantendo-se todas as outras variáveis constantes.

- **Geomorfologia-Planícies (Aratu)**

A chance do sítio ser Aratu em relação a ser Tupiguarani com Geomorfologia igual a Planície é 46% (i.e., $1 - \exp(-0,62) = -0,46$) menor que a chance do sítio ser Aratu em relação a ser Tupiguarani com Geomorfologia igual a referência mantendo-se todas as outras variáveis constantes.

- **Tipo cerâmico pré-colonial-Simples/Dec.Plástica (Aratu)**

A chance do sítio ser Aratu em relação a ser Tupiguarani com Tipo cerâmico pré-colonial igual a Simples/Dec.Plástica é 2531% (i.e., $1 - \exp(-3,27) = 25,31$) maior que a chance do sítio ser Aratu em relação a ser Tupiguarani com Tipo cerâmico pré-colonial igual a referência mantendo-se todas as outras variáveis constantes.

- **Tipo cerâmico pré-colonial-Todos/Não aplicável (Aratu)**

A chance do sítio ser Aratu em relação a ser Tupiguarani com Tipo cerâmico pré-colonial igual a Todos/Não aplicável é 90% (i.e., $1 - \exp(-2,33) = 0,90$) menor que a chance do sítio ser Aratu em relação a ser Tupiguarani com Tipo cerâmico pré-colonial igual a referência mantendo-se todas as outras variáveis constantes.

- **Tipo cerâmico pré-colonial-Não especificado (Aratu)**

A chance do sítio ser Aratu em relação a ser Tupiguarani com Tipo cerâmico pré-colonial igual a Não especificado é 120% (i.e., $1 - \exp(0,79) = 1,20$) maior que a chance do sítio ser Aratu em relação a ser Tupiguarani com Tipo cerâmico pré-colonial igual a referência mantendo-se todas as outras variáveis constantes.

- **Material cultural-Cerâmica/Outro (Aratu)**

A chance do sítio ser Aratu em relação a ser Tupiguarani com Material cultural igual a Cerâmica/Outro é 410% maior (i.e., $1 - \exp(1,63) = 4,10$) que a chance do sítio ser Aratu em relação a ser Tupiguarani com Material cultural igual a referência mantendo-se todas as outras variáveis constantes.

- **Intercepto (Una)**

A chance de ser Una em relação a ser Tupiguarani é $\exp(-6,13) = 0,002$ para um sítio com todas as variáveis nos valores e níveis de referência, ou seja, a probabilidade de ser Una é 99% (pois, $1 - \exp(-6,13) = 0,99$) menor que a probabilidade de ser Tupiguarani para esse sítio.

- **Distância padronizada (Una)**

A chance do sítio ser Una em relação a ser Tupiguarani com acréscimo em 1 unidade na Distância padronizada é 80% (i.e., $1 - \exp(-1,59) = -0,80$) menor que a chance do sítio ser Una em relação a ser Tupiguarani sem esse acréscimo na Distância padronizada mantendo-se todas as outras variáveis constantes.

- **Coordenada E padronizada (Una)**

A chance do sítio ser Una em relação a ser Tupiguarani com acréscimo em 1 unidade na Coordenada E padronizada é 79% (i.e., $1 - \exp(-1,57) = -0,79$) menor que a chance do sítio ser Una em relação a ser Tupiguarani sem esse acréscimo na Coordenada E padronizada mantendo-se todas as outras variáveis constantes.

- **Coordenada N padronizada (Una)**

A chance do sítio ser Una em relação a ser Tupiguarani com acréscimo em 1 unidade na Coordenada N padronizada é 74% (i.e., $1 - \exp(-1,35) = -0,74$) menor que a chance do sítio ser Una em relação a ser Tupiguarani sem esse acréscimo na Coordenada N padronizada mantendo-se todas as outras variáveis constantes.

- **Geomorfologia-Planícies (Una)**

A chance do sítio ser Una em relação a ser Tupiguarani com Geomorfologia igual a Planície é 6180% (i.e., $1 - \exp(4,14) = 61,80$) maior que a chance do

sítio ser Una em relação a ser Tupiguarani com Geomorfologia igual a referência mantendo-se todas as outras variáveis constantes.

- **Tipo cerâmico pré-colonial-Simples/Dec.Plástica (Una)**

A chance do sítio ser Una em relação a ser Tupiguarani com Tipo cerâmico pré-colonial igual a Simples/Dec.Plástica é 16139% (i.e., $1 - \exp(5,09) = 161,32$) maior que a chance do sítio ser Una em relação a ser Tupiguarani com Tipo cerâmico pré-colonial igual a referência mantendo-se todas as outras variáveis constantes.

- **Tipo cerâmico pré-colonial-Todos/Não aplicável (Una)**

A chance do sítio ser Una em relação a ser Tupiguarani com Tipo cerâmico pré-colonial igual a Todos/Não aplicável é 58% (i.e., $1 - \exp(-0,88) = -0,58$) menor que a chance do sítio ser Una em relação a ser Tupiguarani com Tipo cerâmico pré-colonial igual a referência mantendo-se todas as outras variáveis constantes.

- **Tipo cerâmico pré-colonial-Não especificado (Una)**

A chance do sítio ser Una em relação a ser Tupiguarani com Tipo cerâmico pré-colonial igual a Não especificado é 65% (i.e., $1 - \exp(0,50) = 0,65$) maior que a chance do sítio ser Una em relação a ser Tupiguarani com Tipo cerâmico pré-colonial igual a referência mantendo-se todas as outras variáveis constantes

- **Material cultural-Cerâmica/Outro (Una)**

A chance do sítio ser Una em relação a ser Tupiguarani com Material cultural igual a Cerâmica/Outro é 8% (i.e., $1 - \exp(-0,08) = -0,08$) menor que a chance do sítio ser Una em relação a ser Tupiguarani com Material cultural igual a referência mantendo-se todas as outras variáveis constantes.

Estudamos os resíduos dos modelos para verificar o ajuste do modelo. Os resultados se mostraram satisfatórios indicando que o modelo está bem ajustado. Para mais detalhes sobre essas verificações, veja o Apêndice D.

Verificamos a qualidade do modelo observando o número de acertos na matriz de confusão (Tabela 7). Para tal, definimos a regra de classificação, com o auxílio pesquisadora, por meio da probabilidade de o sítio pertencer a cada filiação. O Quadro 2 apresenta a regra de classificação sugerida aos sítios Não Cerâmicos. Os erros de classificações entre as filiações têm o mesmo peso. Sendo assim, maximizamos a diagonal principal da matriz contida na Tabela 7, apresentamos os resultados para a pesquisadora e em conjunto desenvolvemos as regras contidas no Quadro 2.

Quadro 2. Método de classificação sugerido dos sítios Cerâmicos, segundo o modelo trinomial

Filiação Cultural	Regra de classificação
Una	Probabilidade de o sítio ser Una, segundo o modelo trinomial, maior que 40%.
Aratu	Não ser classificado como Una e ter probabilidade de o sítio ser Aratu, segundo o modelo trinomial, maior que 25%.
Tupiguarani	Não ser classificado como Una ou Aratu.

Tabela 7. Matriz de confusão para o modelo final para não ceramistas

		PREDITO		
		Tupiguarani	Aratu	Una
REAL	Tupiguarani	79	12	1
	Aratu	6	22	1
	Una	3	1	10

Uma questão pertinente levantada pela pesquisadora foi a necessidade de se alterar as classificações dessas filiações culturais, como a não distinção entre Aratu e Una na arqueologia do estado de São Paulo, por exemplo. Como a Tabela 7 mostra, a quantidade de erros entre Aratu e Una é baixa (um sítio Aratu como Una e um sítio Una

como Aratu). Logo, podemos concluir que há uma distinção intrínseca na modelagem e classificação entre as duas filiações, já que há pouca confusão entre elas.

Segundo a regra do Quadro 2, fizemos um mapa com a classificação dos sítios sem filiação definida anteriormente. Utilizamos apenas os sítios Não Definido cuja classificação é não ceramista. A Figura 19 apresenta o mapa em questão.

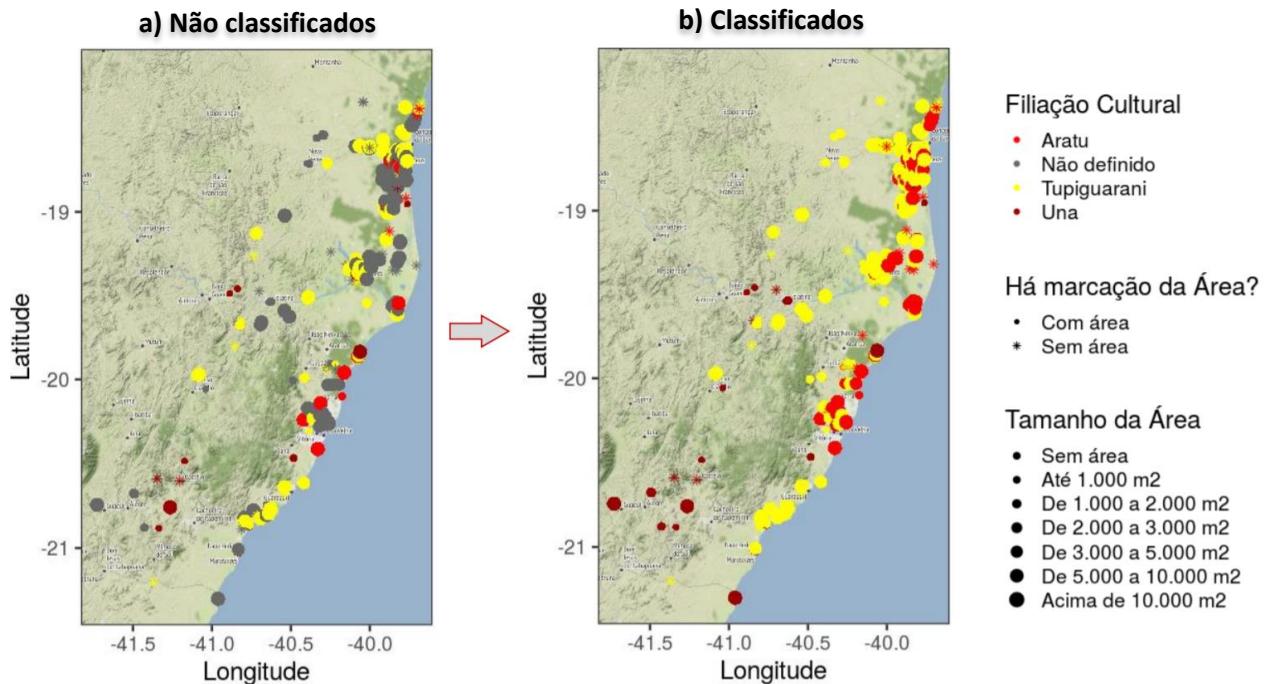


Figura 19. Classificação dos sítios Não Definido

6.3 Classificação dos sítios

Os métodos de modelagem anteriormente descritos (e também no Apêndice C) indicam um modelo estatístico para obtenção, dentre outras coisas, das probabilidades de o sítio pertencer a cada filiação cultural. Com isso, recomendamos o uso dos modelos para classificação das filiações culturais dos sítios arqueológicos da seguinte maneira:

- O pesquisador deve decidir se o sítio é Cerâmico ou não;
- Se for um sítio Cerâmico, calcular as probabilidades de o sítio ser Aratu, Tupiguarani ou Una por meio do modelo trinomial estimado (Apêndice C) e aplicar as nossas regras apresentadas no Quadro 2;
- Se for um sítio Não Cerâmico, calcular as probabilidades de o sítio ser Itaipu ou Sambaqui e classificar o sítio de acordo com o prescrito pelo Quadro 1;
- Caso não seja possível classificar o sítio entre Cerâmico e Não Cerâmico, recomendamos que o pesquisador calcule as probabilidades para cada um dos modelos (trinomial e binomial), classifique-os conforme instruímos anteriormente e avalie qual das classificações (Cerâmica ou Não Cerâmico) é a mais pertinente de acordo com outros critérios arqueológicos para aquele sítio em questão.

Para facilitar esse procedimento de classificação, desenvolvemos uma planilha dinâmica que está disponível na pasta do projeto (Seção OUTROS). As Figuras 20, 21 e 22 apresentam o funcionamento da planilha para um sítio Cerâmico, Não Cerâmico e indefinido, respectivamente. Além de fornecer as variáveis necessárias para classificação, o usuário pode alterar as regras para a classificação dos sítios.

CLASSIFICADOR DE SÍTIOS CAPIXABAS																										
<table border="1"> <thead> <tr> <th colspan="2">CERÂMICO OU NÃO CERÂMICO</th> <th colspan="2">COORDENADAS DE LOCALIZAÇÃO</th> <th colspan="2">DISTÂNCIA PARA O RECURSO HÍDRICO</th> </tr> </thead> <tbody> <tr> <td>Como classificar esse sítio?</td> <td>Cerâmico</td> <td><input checked="" type="checkbox"/> Coordenada E (UTM) 400000</td> <td>Coordenada N (UTM) 7800000</td> <td>Qual a Distância (em metros)?</td> <td>0.5</td> </tr> <tr> <td colspan="2"></td> <td colspan="2">GEOMORFOLOGIA</td> <td colspan="2">TIPO CERÂMICO</td> </tr> <tr> <td colspan="2"></td> <td colspan="2">Qual a Geomorfologia? Planícies</td> <td colspan="2">Qual o Tipo Cerâmico? Simples/Dec. Plástica</td> </tr> </tbody> </table>			CERÂMICO OU NÃO CERÂMICO		COORDENADAS DE LOCALIZAÇÃO		DISTÂNCIA PARA O RECURSO HÍDRICO		Como classificar esse sítio?	Cerâmico	<input checked="" type="checkbox"/> Coordenada E (UTM) 400000	Coordenada N (UTM) 7800000	Qual a Distância (em metros)?	0.5			GEOMORFOLOGIA		TIPO CERÂMICO				Qual a Geomorfologia? Planícies		Qual o Tipo Cerâmico? Simples/Dec. Plástica	
CERÂMICO OU NÃO CERÂMICO		COORDENADAS DE LOCALIZAÇÃO		DISTÂNCIA PARA O RECURSO HÍDRICO																						
Como classificar esse sítio?	Cerâmico	<input checked="" type="checkbox"/> Coordenada E (UTM) 400000	Coordenada N (UTM) 7800000	Qual a Distância (em metros)?	0.5																					
		GEOMORFOLOGIA		TIPO CERÂMICO																						
		Qual a Geomorfologia? Planícies		Qual o Tipo Cerâmico? Simples/Dec. Plástica																						
<table border="1"> <thead> <tr> <th colspan="2">PROBABILIDADES - CERÂMICO</th> <th colspan="2">CORTES PARA O SÍTIO CERÂMICO</th> <th colspan="2">CLASSIFICAÇÃO - CERÂMICO</th> </tr> </thead> <tbody> <tr> <td>Probabilidade de ser Aratu</td> <td>18,912%</td> <td>Probabilidade mínima para ser Una</td> <td>40,000%</td> <td>O Sítio Cerâmico é:</td> <td>Una</td> </tr> <tr> <td>Probabilidade de ser Tupiguarani</td> <td>3,071%</td> <td>Não sendo Una, Probabilidade mínima para ser Aratu</td> <td>25,000%</td> <td></td> <td></td> </tr> <tr> <td>Probabilidade de ser Una</td> <td>78,016%</td> <td></td> <td></td> <td></td> <td></td> </tr> </tbody> </table>			PROBABILIDADES - CERÂMICO		CORTES PARA O SÍTIO CERÂMICO		CLASSIFICAÇÃO - CERÂMICO		Probabilidade de ser Aratu	18,912%	Probabilidade mínima para ser Una	40,000%	O Sítio Cerâmico é:	Una	Probabilidade de ser Tupiguarani	3,071%	Não sendo Una, Probabilidade mínima para ser Aratu	25,000%			Probabilidade de ser Una	78,016%				
PROBABILIDADES - CERÂMICO		CORTES PARA O SÍTIO CERÂMICO		CLASSIFICAÇÃO - CERÂMICO																						
Probabilidade de ser Aratu	18,912%	Probabilidade mínima para ser Una	40,000%	O Sítio Cerâmico é:	Una																					
Probabilidade de ser Tupiguarani	3,071%	Não sendo Una, Probabilidade mínima para ser Aratu	25,000%																							
Probabilidade de ser Una	78,016%																									

Figura 20. Planilha de classificação para um sítio Cerâmico

CLASSIFICADOR DE SÍTIOS CAPIXABAS					
CERÂMICO OU NÃO CERÂMICO		COORDENADAS DE LOCALIZAÇÃO		DISTÂNCIA PARA O RECURSO HÍDRICO	
Como classificar esse sítio?	Não Cerâmico	Coordenada E (UTM)	400000	Coordenada N (UTM)	7800000
ELEVAÇÃO		Qual a Distância (em metros)?		0,5	
Qual a Elevação (em metros)?	100	Probabilidade mínima para ser Sambaqui	80,000%	O Sítio Não Cerâmico é:	Sambaqui
PROBABILIDADES - NÃO CERÂMICO		CORTES PARA O SÍTIO NÃO CERÂMICO		CLASSIFICAÇÃO - NÃO CERÂMICO	
Probabilidade de ser Itaipu	1,953%	Probabilidade mínima para ser Sambaqui	80,000%	O Sítio Não Cerâmico é:	Sambaqui
Probabilidade de ser Sambaqui	98,047%				

Figura 21. Planilha de classificação para um sítio Não Cerâmico

CLASSIFICADOR DE SÍTIOS CAPIXABAS					
CERÂMICO OU NÃO CERÂMICO		COORDENADAS DE LOCALIZAÇÃO		DISTÂNCIA PARA O RECURSO HÍDRICO	
Como classificar esse sítio?	Sem classificação	Coordenada E (UTM)	400000	Coordenada N (UTM)	7800000
ELEVAÇÃO		GEOMORFOLOGIA		TIPO CERÂMICO	
Qual a Elevação (em metros)?	100	Qual a Geomorfologia?	Planícies	Qual o Tipo Cerâmico?	Simples/Dec. Plástica
PROBABILIDADES - NÃO CERÂMICO		CORTES PARA O SÍTIO NÃO CERÂMICO		CLASSIFICAÇÃO - NÃO CERÂMICO	
Probabilidade de ser Itaipu	1,953%	Probabilidade mínima para ser Sambaqui	80,000%	O Sítio Não Cerâmico é:	Sambaqui
Probabilidade de ser Sambaqui	98,047%				
PROBABILIDADES - CERÂMICO		CORTES PARA O SÍTIO CERÂMICO		CLASSIFICAÇÃO - CERÂMICO	
Probabilidade de ser Aratu	56,246%	Probabilidade mínima para ser Una	40,000%	O Sítio Cerâmico é:	Una
Probabilidade de ser Tupiguarani	1,790%	Não sendo Una, Probabilidade mínima para ser Aratu	25,000%		
Probabilidade de ser Una	41,965%				

Figura 22. Planilha de classificação para um sítio não classificado como Cerâmico ou Não Cerâmico

6.4 Estudo de caso para a ocupação da tradição Aratu em Linhares

Desenvolvemos também um estudo de caso para a filiação cultural Aratu no município de Linhares (Figura 23). Esse município possui muitas lagoas, tornando seus recursos hídricos abundantes. A pesquisadora defende que há uma ocupação histórica do município por parte das tribos da Tradição Aratu, enquanto outros arqueólogos defendem que os sítios Não Definido são da Tradição Tupiguarani. Essa é a motivação do estudo de caso a seguir.

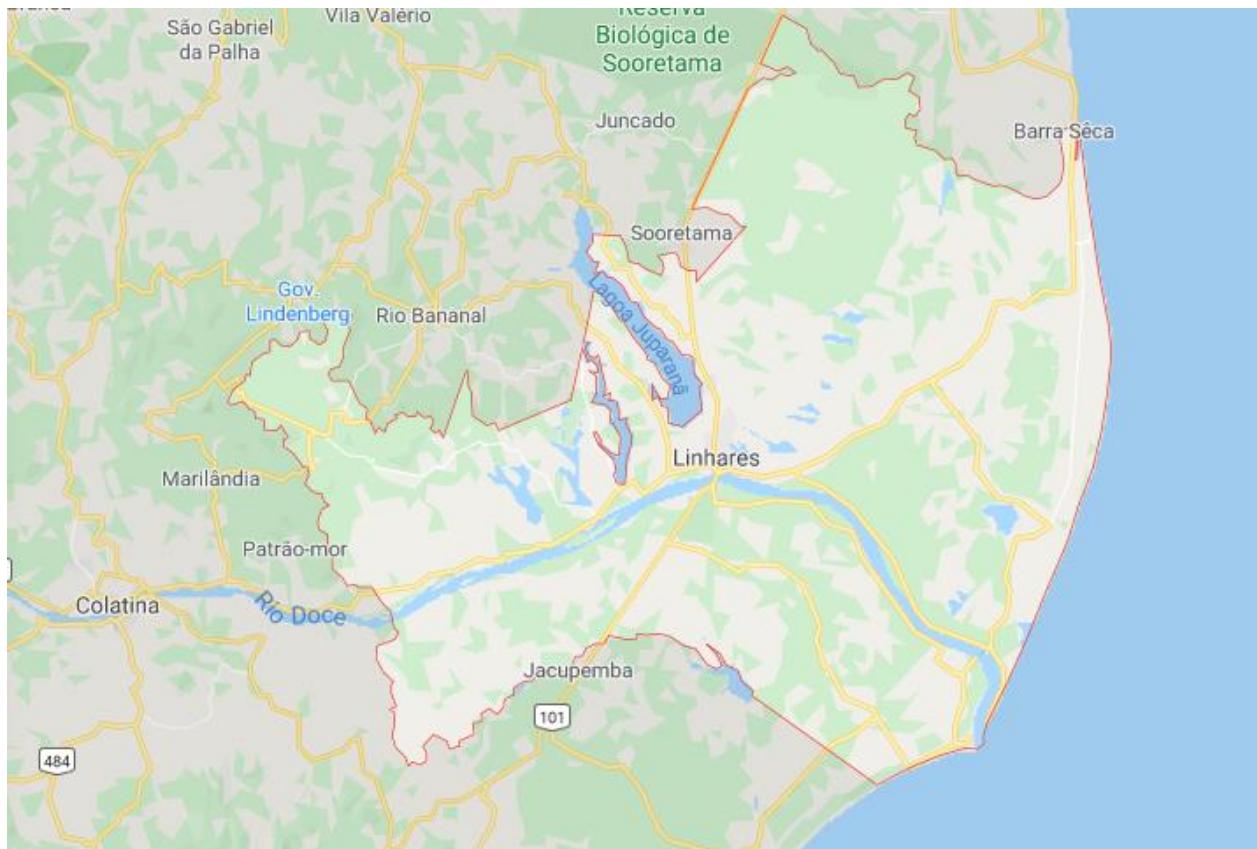


Figura 23. Mapa do município de Linhares

Inicialmente, verificamos a porcentagem de classificação de sítios Aratu em todo o estado do Espírito Santo (28%) frente a mesma classificação apenas nos sítios localizados em Linhares (37%).

Posteriormente, calculamos as probabilidades das filiações Cerâmicas com o modelo trinomial em todo município de Linhares. Para isso, assumimos que os sítios classificados são Cerâmicos, possuem **Distância** de algum recurso hídrico igual a média dos sítios de Linhares (1,29 km), estão em Planícies (categoria associada a variável Geomorfologia) e variamos os Tipos cerâmicos entre “Simples/Dec.Pintada” (primeira linha da Figura 24) e “Simples/Dec.Plástica” (segunda linha da Figura 24). A Figura 24 mostra as probabilidades de cada uma das filiações cerâmicas em cada ponto do município.

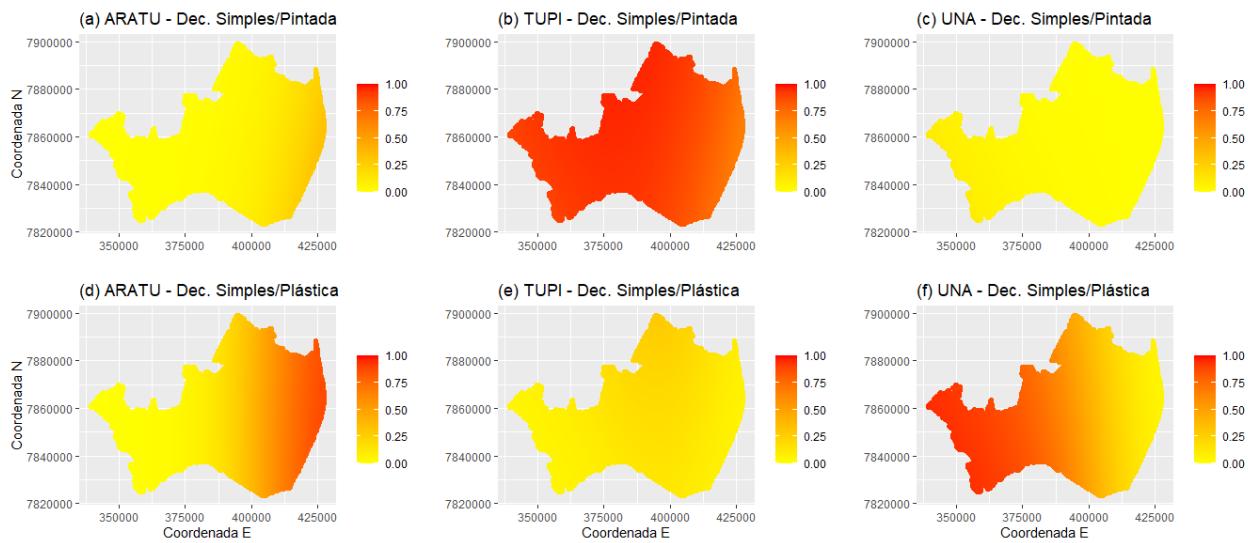


Figura 24. Probabilidades de classificação para Linhares

Pela Figura 24, temos que o gráfico (a) ilustra a probabilidade de ter filiação cultural igual a Aratu para todo o município de Linhares quando a Distância é 1,29 km (média de Linhares), a Geologia é Planície, o Tipo cerâmico é Simples/Dec.Pintada e o Material cultural é Cerâmica. Os gráficos (b) e (c) mostram as probabilidades de ser Tupiguarani e Una respectivamente, para as mesmas condições de (a). O gráfico (d) mostra a probabilidade de ter filiação cultural igual a Aratu para todo o município de Linhares nas mesmas condições anteriores, mas agora mudando o Tipo cerâmico para Simples/Dec.Plástica. Os gráficos (e) e (f) mostram as probabilidades de ser Tupiguarani e Una respectivamente, para as mesmas condições de (d).

Podemos inferir a partir dos mapas da Figura 24 que os sítios com essas características em Linhares têm uma probabilidade maior de serem Aratu do que Tupiguarani ou Una. As relações vistas explicitam a predominância da tradição Aratu no litoral de Linhares, enquanto no interior há uma predominância de Tupiguarani para o Tipo cerâmico “Simples/Dec.Pintada” e uma divisão entre as filiações para o Tipo cerâmico “Simples/Dec.Plástica”.

Também desenvolvemos GIFs animados disponíveis na pasta do projeto (Seção OUTROS). A partir deles é possível notar que Aratu é muito mais predominante no litoral.

Contudo, conforme a Distância para a água aumenta, quando o tipo cerâmico é “Simples/Dec.Pintada”, há uma predominância de Tupiguarani em todo o município de Linhares.

Assim, entendemos que a região de Linhares possui características que aumentam a probabilidade da filiação Aratu, o que corrobora a interpretação histórica dos deslocamentos dessas tribos para essa região lagunar.

7. Conclusões

Durante o desenvolvimento do projeto foi possível responder a todas as perguntas da pesquisadora, bem como, manter um certo rigor estatístico. Os conhecimentos arqueológicos prévios da pesquisadora foram ratificados na Análise descritiva.

Através das classificações propostas neste relatório, é possível entender e interpretar cada uma das informações e quantificar suas relevâncias. Uma das vantagens dessa abordagem consiste na escrita das relações por meio de chances e razões de chances, pois conseguimos associar determinadas preferências ou características para determinadas filiações. Por exemplo, a chance do sítio ser Aratu em relação a ser Tupiguarani com Tipo cerâmico pré-colonial igual a Simples/Dec. Plástica é 2531% maior que a chance do sítio ser Aratu em relação a ser Tupiguarani com Tipo cerâmico pré-colonial igual a Simples/Dec. Pintada, mantendo-se todas as outras variáveis constantes, o que já era esperado na arqueologia, por conta do tipo de cerâmica feito por cada cada filiação, mas que agora pôde ser quantificado.

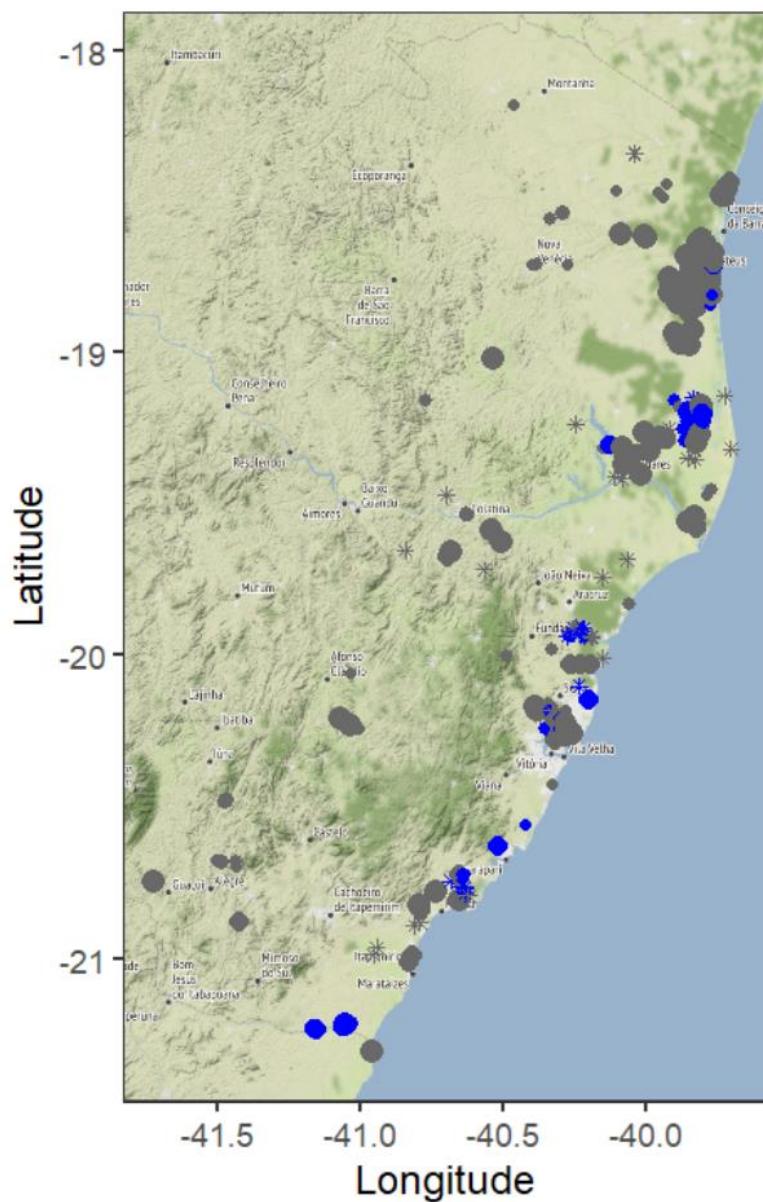
Além disso, a partir desse sistema de modelagem, desenvolvemos a planilha de classificação para que o pesquisador identifique a filiação mais provável para determinado sítio arqueológico.

Uma abordagem alternativa seria desenvolver a modelagem via Árvore de Decisão, por exemplo. Os métodos devem produzir resultados semelhantes, pois as Análises descritiva e inferencial estão em sintonia.

Uma possível continuação para o estudo seria complementar o conjunto de dados para a inclusão de outras informações arqueológicas e o desenvolvimento de um guia inteligente de classificação (semelhante ao mapa feito no estudo de caso de Linhares) para todo o estado do Espírito Santo.

APÊNDICE A

Figuras



Filiação Cultural

- Não definido
 - Sambaqui

Há marcação da Área?

- Com área
 - * Sem área

Tamanho da Área

- Sem área
 - Até 1.000 m²
 - De 1.000 a 2.000 m²
 - De 2.000 a 3.000 m²
 - De 3.000 a 5.000 m²
 - De 5.000 a 10.000 m²
 - Acima de 10.000 m²

Figura A.1 Distribuição da filiação Sambaqui com sítios classificados como Não Definido de acordo com a localização

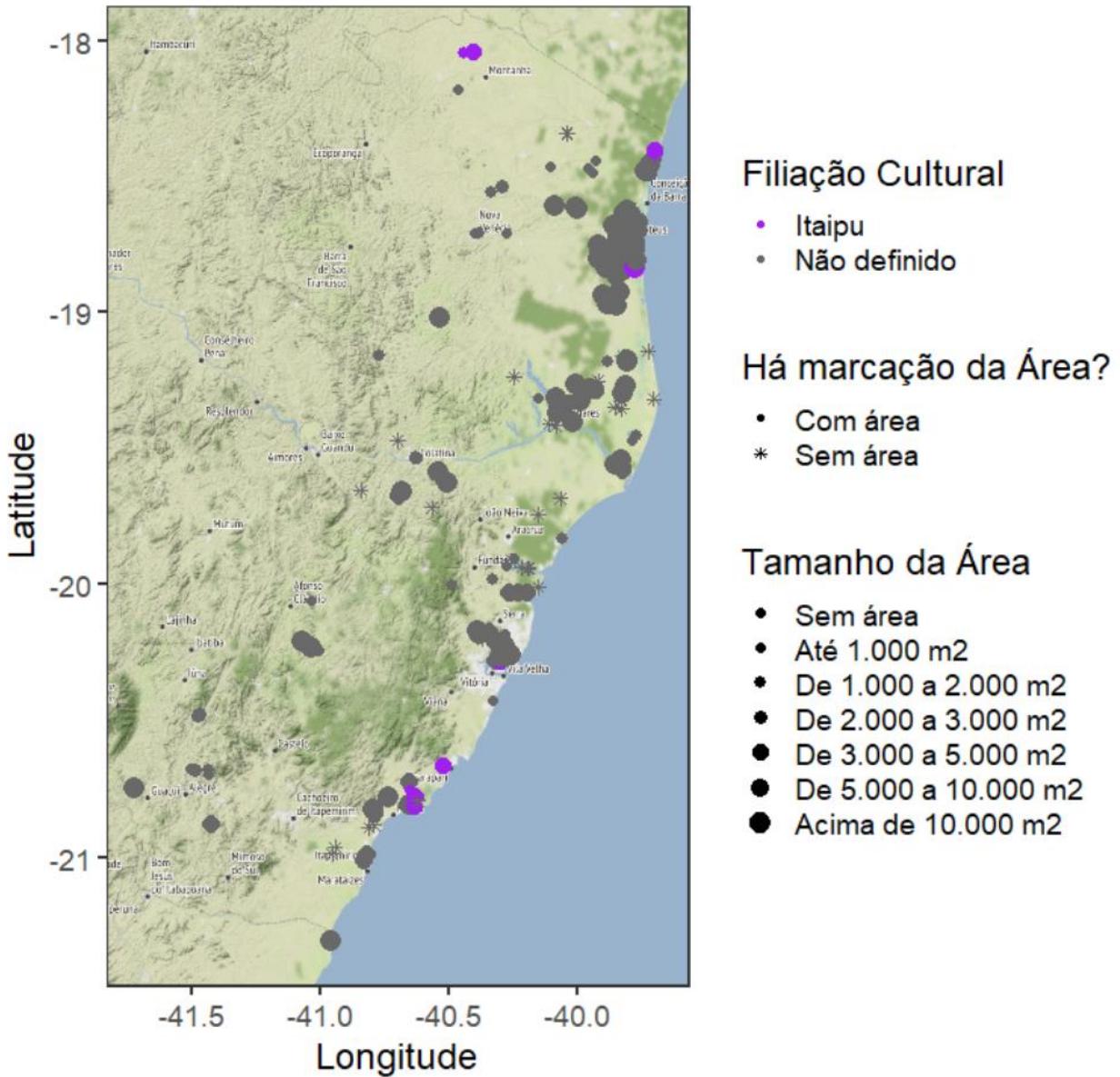


Figura A.2 Distribuição da filiação Itaipu com sítios classificados como Não Definido de acordo com a localização

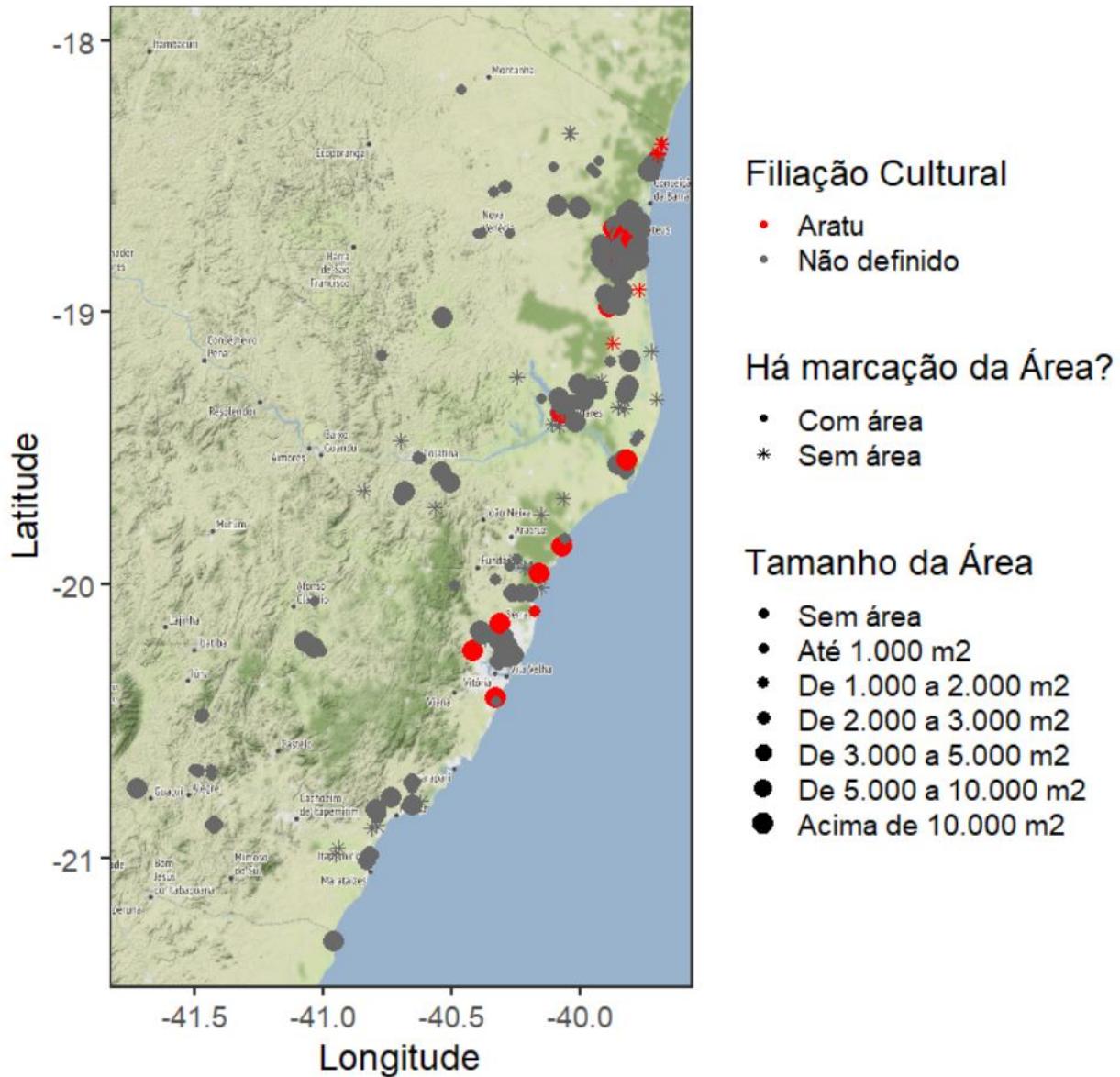


Figura A.3 Distribuição da filiação Aratu com sítios classificados como Não Definido de acordo com a localização

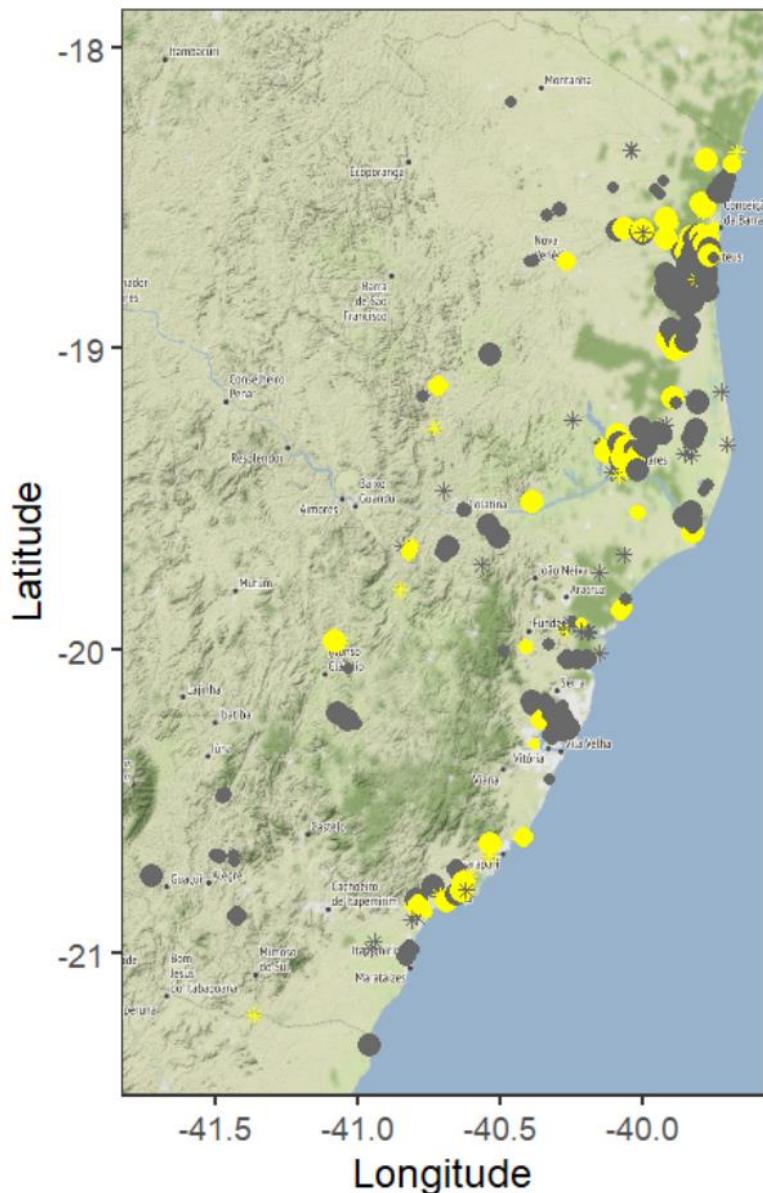


Figura A.4 Distribuição da filiação Tupiguarani com sítios classificados como Não Definido de acordo com a localização

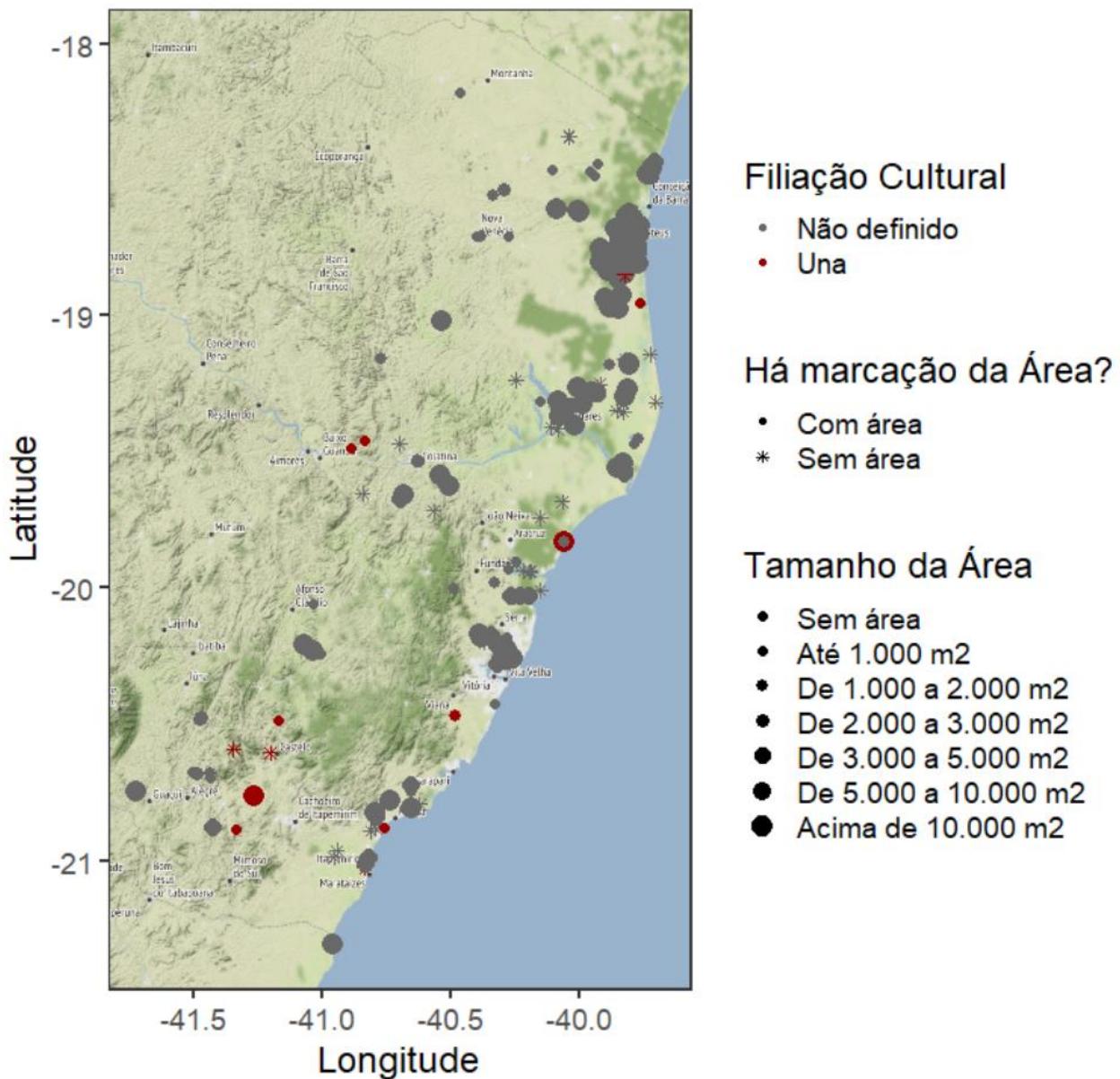
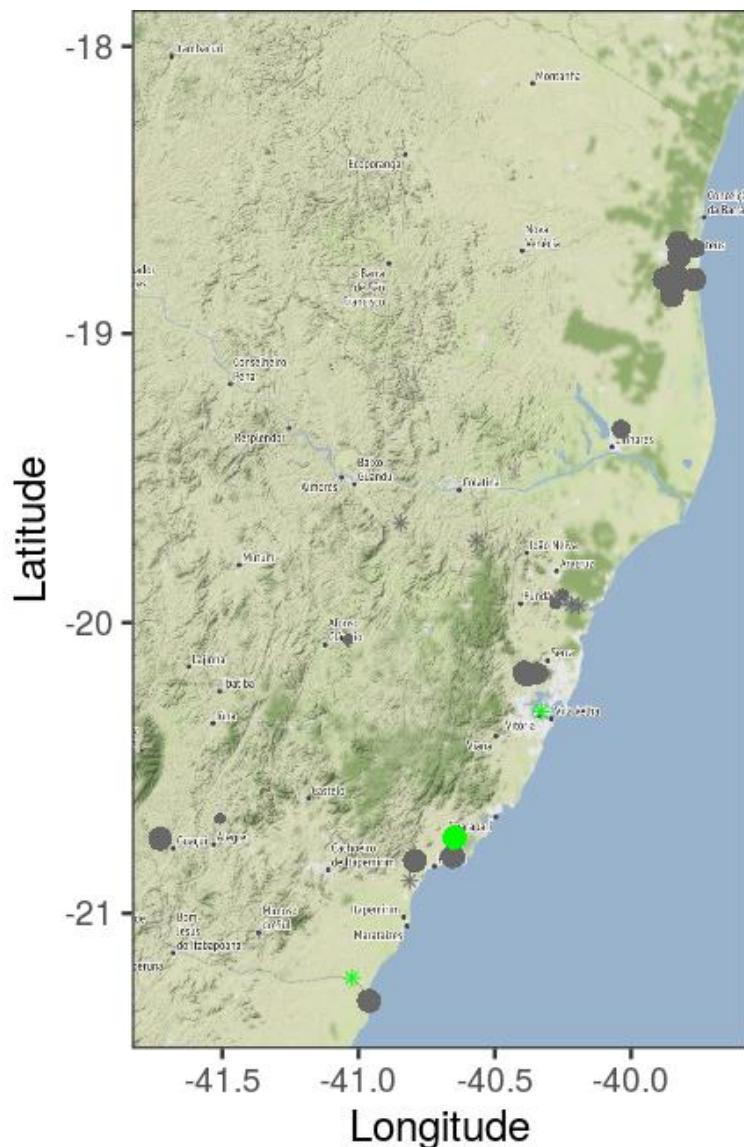


Figura A.5 Distribuição da filiação Una com sítios classificados como Não Definido de acordo com a localização



Filiação Cultural

- Não definido
- Sambaqui/Tupiguarani

Há marcação da Área?

- Com área
- * Sem área

Tamanho da Área

- Sem área
- Até 1.000 m²
- De 1.000 a 2.000 m²
- De 2.000 a 5.000 m²
- De 5.000 a 10.000 m²
- Acima de 10.000 m²

Figura A.6 Distribuição da filiação Sambaqui/Tupiguarani com sítios classificados como Não Definido de acordo com a localização

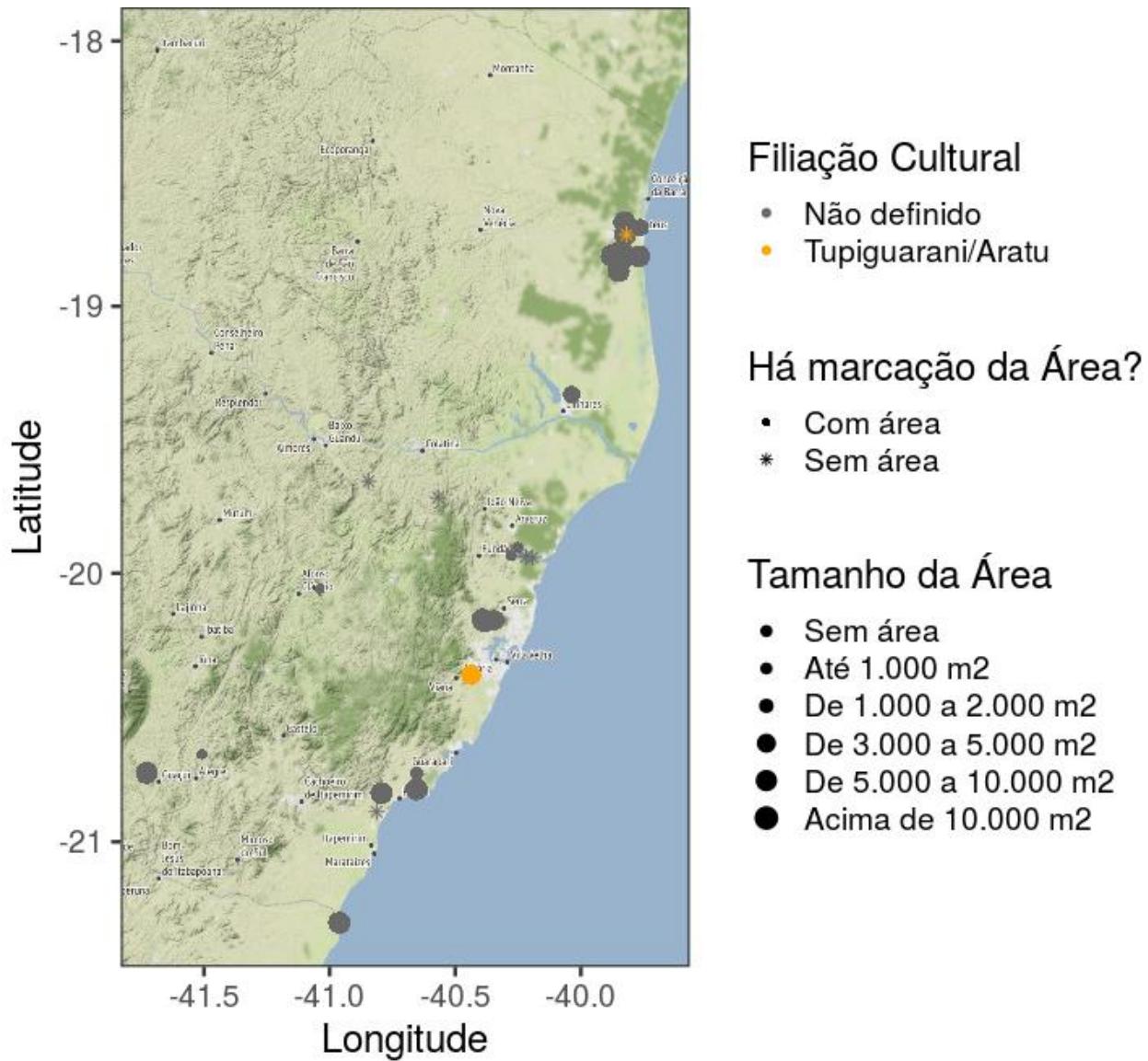


Figura A.7 Distribuição da filiação Tupiguarani/Aratu com sítios classificados como Não Definido de acordo com a localização

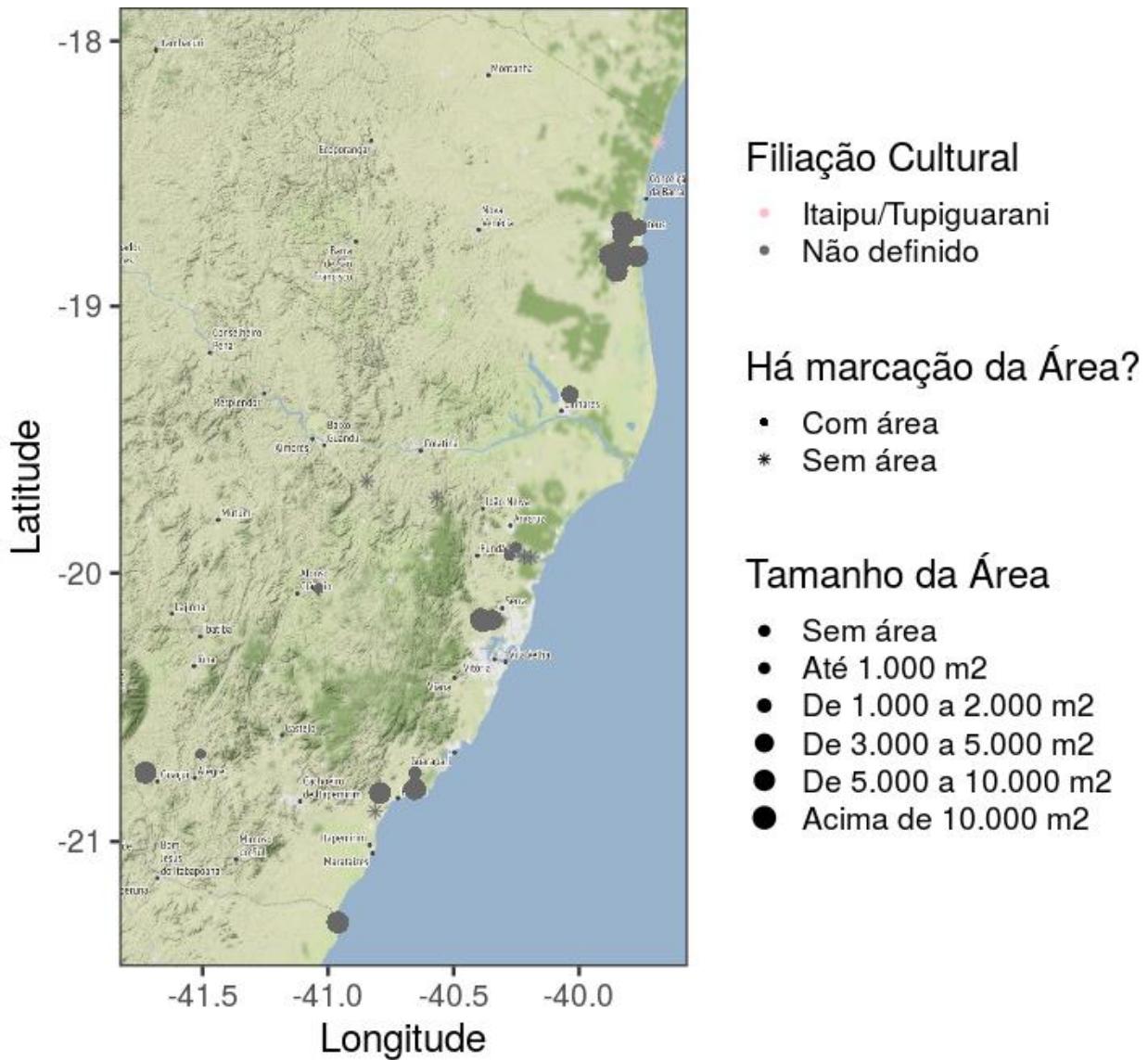


Figura A.8 Distribuição da filiação Itaipu/Tupiguarani com sítios classificados como Não Definido de acordo com a localização

APÊNDICE B

Tabelas

Depois da Análise descritiva, as variáveis excluídas pela pesquisadora foram: **Área, Clima, Unidade, Uso, Vegetação e Geologia**. As variáveis padronizadas foram: **Distância, Coordenada E e Coordenada N**. Padronizamos as variáveis subtraindo a média dos sítios com filiação definida e dividindo pelo desvio-padrão também desses sítios. Na Tabela B.1 apresentamos esses valores utilizados na padronização.

Tabela B.1 Valores de padronização utilizados nos modelos binomial e trinomial

	Média	Desvio-padrão
Distância	0,73	0,87
Coordenada E	375019,70	44324,03
Coordenada N	7839094,00	89423,26

As variáveis restantes foram todas categóricas, então para as variáveis com mais de dois níveis juntamos esses níveis, ficando assim com o número de categorias reduzidas. Nas Tabelas B.2 a B.5 apresentamos como foram feitas essas junções para cada variável.

Tabela B.2 Novas categorias correspondentes para **Material cultural**

Antigas categorias	Novas categorias
CP	Cerâmica
Lítico	Outro
Malacológico	Outro
CP/Lítico	Cerâmica/Outro
CP/CC	Cerâmica
CP/Lítico/Malacológico/CC	Cerâmica/Outro
CP/Lítico	Cerâmica/Outro
Lítico/Malacológico/Ósseo	Outro
Lítico/CC	Cerâmica/Outro
Todos	Cerâmica/Outro
CC/Malacológico/Ósseo/Lítico	Cerâmica/Outro
CP/Lítico/Ósseo/Malacológico	Cerâmica/Outro

Tabela B.3 Novas categorias correspondentes para **Tipo cerâmico pré-colonial**

Antigas categorias	Novas categorias
Simples	Simples/Dec.Pintada
Decoração Pintada	Simples/Dec.Pintada
Decoração Plástica	Simples/Dec.Plástica
Simples / Dec. Plástica	Simples/Dec.Plástica
Simples / Dec. Pintada	Simples/Dec.Pintada
Todos	Todos/Não aplicável
Não especificado	Não especificado
Não se aplica	Todos/Não aplicável

Tabela B.4 Novas categorias correspondentes para **Geomorfologia**

Antigas categorias	Novas categorias
Chapadas/Planaltos/Patamares	Chapadas/Planaltos/Patamares
Depressão	Tabuleiros/Depressão/Escarpas/Reversos
Escarpas/Reversos	Tabuleiros/Depressão/Escarpas/Reversos
Planícies(Fluviais/Fluviolacustres)	Planícies
Planícies(Marinhas/Fluviomarinhas/ Fluviolacustres)	Planícies
Tabuleiros Costeiros	Tabuleiros/Depressão/Escarpas/Reversos

Tabela B.5 Novas categorias correspondentes para **Pedologia**

Antigas categorias	Novas categorias
Latossolos	Latossolos
Argilossolos	Argilossolos/Neossolos
Neossolos	Argilossolos/Neossolos
Argissolos	Argilossolos/Neossolos
Chernossolos	Chernossolos/Gleissolos
Gleissolos	Chernossolos/Gleissolos
Espodossolos	Outro
Cambissolo	Outro
Não informado	Outro
Solos Halomórficos	Outro

Com a junção dos níveis dessas variáveis agora temos observações em todos os níveis de todas as variáveis considerando a divisão por filiações ceramistas e não ceramistas. Nas Tabelas B.6 a B.9 apresentamos as tabelas de frequências para as variáveis que tiveram seus níveis juntados.

Tabela B.6 Tabela de frequências para Material cultural

Filiação cultural	Material cultural						Total	
	Cerâmica		Cerâmica/Outro		Outro			
	n	%	n	%	n	%		
Itaipu	0	0	3	21,42	11	78,57	14	100
Sambaqui	0	0	9	9,37	87	90,62	96	100
Aratu	10	34,48	19	65,51	0	0	29	100
Tupiguarani	50	54,34	42	45,65	0	0	92	100
Una	6	42,85	8	57,14	0	0	14	100

Tabela B.7 Tabela de frequências para Geomorfologia

Filiação cultural	Geomorfologia						Total	
	Tabuleiros/Depressão/ Escarpas/Reversos			Planícies				
	n	%	n	%	n	%		
Itaipu	11	78,57	3	21,42	14	100		
Sambaqui	68	70,83	28	29,16	96	100		
Aratu	11	37,93	18	62,06	29	100		
Tupiguarani	54	58,69	38	41,30	92	100		
Una	10	71,42	4	28,57	14	100		

Tabela B.8 Tabela de frequências para Pedologia

Filiação cultural	Pedologia						Total			
	Outro		Latossolos		Argilossolos/ Neossolos					
	n	%	n	%	n	%				
Itaipu	2	14,28	1	7,14	5	35,71	6	42,85	14	100
Sambaqui	15	15,62	6	6,25	36	37,50	39	40,62	96	100
Aratu	2	6,89	2	6,89	17	58,62	8	27,58	29	100
Tupiguarani	8	8,69	11	11,95	62	67,39	11	11,95	92	100
Una	1	7,14	3	21,42	8	57,14	2	14,28	14	100

Tabela B.9 Tabela de frequências para **Tipo cerâmico pré-colonial**

Filiação cultural	Tipo cerâmico pré-colonial						Total			
	Simples/Dec. Pintada		Simples/Dec. Plástica		Todos/Não aplicável					
	n	%	n	%	n	%	n	%		
Aratu	2	6,89	11	37,93	1	3,44	15	51,72	29	100
Tupiguarani	21	22,82	3	3,26	29	31,52	39	42,59	92	100
Una	3	21,42	2	14,28	1	7,14	8	57,14	14	100

APÊNDICE C

Metodologia

C.1. Considerações iniciais

Podemos considerar que todas as filiações culturais capixabas estão contempladas no estudo Sambaqui e Itaipu (Não Cerâmicas) e Aratu, Tupiguarani e Una (Cerâmicas).

Portanto, um sítio “Não Definido” é necessariamente uma das filiações e pertencer a qualquer uma delas é um evento mutuamente exclusivo.

Além disso, temos indícios (variável de Material cultural, entre outras informações) para classificar um sítio entre Não Cerâmico e Cerâmico. Com a sua expertise, o pesquisador também consegue fazer esse tipo de classificação analisando os dados.

Logo, é razoável propor um modelo para cada um desses grupos de filiações culturais. Os modelos propostos foram desenvolvidos com a ideia do esquema presente na Figura C.1:

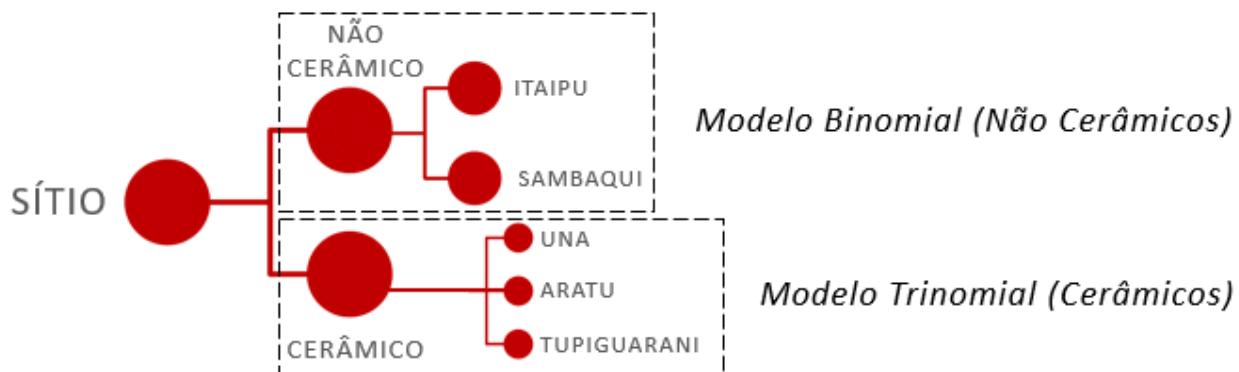


Figura C.1 Esquema de uso dos modelos binomial e trinomial

- A ideia de classificação dos sítios é usar dois modelos utilizando a modelagem Multinomial proposta por Paulino e Singer (2006), sendo:
 - **Modelo binomial:** para os sítios Não Cerâmicos;
 - **Modelo trinomial:** para os sítios Cerâmicos.

- Se não soubermos que um sítio é Cerâmico ou Não Cerâmico e não tivermos indícios de qual dos dois modelos usar para classificar o sítio, aplicaremos os dois modelos para o mesmo sítio, obtendo duas possíveis classificações.
- As variáveis selecionadas para compor o modelo Não Cerâmico foram obtidas dos sítios das filiações Itaipu e Sambaqui. Já as variáveis selecionadas para compor o modelo Cerâmico foram obtidas dos sítios das filiações Aratu, Tupiguarani e Una.
- Utilizamos o pacote *nnet* do R para o desenvolvimento dos modelos.

C.2. Modelo binomial

Para fazermos o modelo binomial, contamos com informações de 110 sítios Não Cerâmicos ($n = 110$) com as filiações Itaipu ($p_{i,1}$) e Sambaqui ($p_{i,2}$).

As variáveis utilizadas no modelo binomial inicial (Não Cerâmico) foram as oito apresentadas na Tabela C.1:

Tabela C.1 Tabela das variáveis iniciais para a construção do modelo binomial

Variável	Tipo	Parametrização
Distância ($k_{i,1}$)	Numérica	$x_{i,1} = (k_{i,1} - \bar{k}_{Dist})/S_{Dist}$
Elevação ($k_{i,2}$)	Numérica	$x_{i,2} = 1(k_{i,2} > 30)$
Coordenada E ($k_{i,3}$)	Numérica	$x_{i,3} = (k_{i,3} - \bar{k}_{Coord\ E})/S_{Coord\ E}$
Coordenada N ($k_{i,4}$)	Numérica	$x_{i,4} = (k_{i,4} - \bar{k}_{Coord\ N})/S_{Coord\ N}$
Pedologia ($k_{i,5}$)	Categórica (4)	$x_{i,5} = 1(k_{i,5} = "Ped_1")$, $x_{i,6} = 1(k_{i,5} = "Ped_2")$, $x_{i,7} = 1(k_{i,5} = "Ped_3")$
Geomorfologia ($k_{i,6}$)	Categórica (2)	$x_{i,8} = 1(k_{i,6} = "Geo_1")$
Período ($k_{i,7}$)	Categórica (2)	$x_{i,9} = 1(k_{i,7} = "Per_1")$
Material Cultural ($k_{i,8}$)	Categórica (2)	$x_{i,10} = 1(k_{i,8} = "MatCult_1")$

Dessa tabela temos que:

- $\bar{k}_{Coord\ E}$ é média da Coordenada E para todos os sítios capixabas, o mesmo vale para a $\bar{k}_{Coord\ N}$ e \bar{k}_{Dist} , sendo que ambas informações são apresentadas na Tabela B.1;
- $S_{Coord\ E}$ é o desvio padrão da Coordenada E para todos os sítios capixabas, o mesmo vale para a $S_{Coord\ N}$ e S_{Dist} , sendo que ambas informações são apresentadas na Tabela B.1;
- $1(k_{i,2} > 30)$ é a função indicadora que vale 1 (um) se a Elevação for maior que 30 e vale 0 (zero) caso contrário;
- $x_{i,5} = 1(k_{i,5} = "Ped_1")$ é a função indicadora que vale 1 (um) se a Pedologia for a 1^a categoria e 0 (zero) caso contrário. O mesmo vale para as outras categorias de Pedologia, bem como as outras variáveis da Tabela C.1. Note que, quando as três funções indicadoras da Pedologia forem zero, significa que ela equivale à 4^a categoria.

Sendo assim, o modelo Não Cerâmico pode ser escrito como:

$$\mathbf{y}_i | \mathbf{x}_i \sim Multinomial(1, \mathbf{p}_i), \quad i = 1, \dots, 110,$$

com,

$$\mathbf{y}_i = (y_{i,1}, y_{i,2}); \quad y_{i,1} + y_{i,2} = 1; \quad \mathbf{y}_i \text{ e } \mathbf{y}_j \text{ indep. para } i \neq j;$$

$$\text{Sítio Itaipu: } \mathbf{y}_i = (1, 0); \quad \text{Sítio Sambaqui: } \mathbf{y}_i = (0, 1);$$

$$\mathbf{x}_i = (x_{i,1}, x_{i,2}, \dots, x_{i,10}); \quad \mathbf{p}_i = (p_{i,1}, p_{i,2}); \quad p_{i,1} + p_{i,2} = 1;$$

$p_{i,1}$: Probabilidade de ser Itaipu; $p_{i,2}$: Probabilidade de ser Sambaqui.

Com a parametrização de logito de referência (Itaipu), temos:

$$\log \left(\frac{p_{i,2}}{p_{i,1}} \right) = \beta_0 + x_{i,1}\beta_1 + x_{i,2}\beta_2 + \dots + x_{i,10}\beta_{10}.$$

Fazendo algumas transformações algébricas, temos que:

$$\log\left(\frac{p_{i,2}}{p_{i,1}}\right) = \beta_0 + x_{i,1}\beta_1 + x_{i,2}\beta_2 + \cdots + x_{i,10}\beta_{10} \Leftrightarrow$$

$$\frac{p_{i,2}}{p_{i,1}} = \exp(\beta_0 + x_{i,1}\beta_1 + x_{i,2}\beta_2 + \cdots + x_{i,10}\beta_{10}) \Leftrightarrow$$

$$p_{i,2} = \exp(\beta_0 + x_{i,1}\beta_1 + x_{i,2}\beta_2 + \cdots + x_{i,10}\beta_{10}) \cdot p_{i,1}.$$

Como $p_{i,1} + p_{i,2} = 1$, então:

$$p_{i,1} = \frac{1}{1 + \exp(\beta_0 + x_{i,1}\beta_1 + x_{i,2}\beta_2 + \cdots + x_{i,10}\beta_{10})}.$$

A escolha desse modelo é pertinente para o problema porque os parâmetros β 's do possuem interpretações:

$$\log\left(\frac{p_{i,2}}{p_{i,1}}\right) = \beta_0 + x_{i,1}\beta_1 + x_{i,2}\beta_2 + \cdots + x_{i,10}\beta_{10},$$

$$\log\left(\frac{p_{i,2}}{p_{i,1}}\right) = \beta_0 + 0 \cdot \beta_1 + 0 \cdot \beta_2 + \cdots + 0 \cdot \beta_{10} = \beta_0.$$

O intercepto (β_0) é o valor do logaritmo da chance de ser Sambaqui para o sítio com as covariáveis numéricas iguais a média do conjunto de dados e as covariáveis categorizadas com o valor de referência.

Aplicando a exponencial no intercepto β_0 , temos o valor da chance:

$$\log\left(\frac{p_{i,2}}{p_{i,1}}\right) = \beta_0 + 0 \cdot \beta_1 + 0 \cdot \beta_2 + \cdots + 0 \cdot \beta_{10} = \beta_0 \Leftrightarrow \frac{p_{i,2}}{p_{i,1}} = \exp(\beta_0)$$

Agora, vamos analisar um dos parâmetros associados às covariáveis. Tomando $\mathbf{x}_{i1} = (1, x_{i,2}, \dots, x_{i,10})$ e $\mathbf{x}_{i0} = (0, x_{i,2}, \dots, x_{i,10})$, temos que:

$$(\mathbf{y}_i, \mathbf{x}_{i1}) \Rightarrow \log\left(\frac{p_{i1,2}}{p_{i1,1}}\right) = \beta_0 + 1 \cdot \beta_1 + \dots + x_{i,10} \cdot \beta_{10} \Leftrightarrow$$

$$\frac{p_{i1,2}}{p_{i1,1}} = \exp(\beta_0 + \beta_1 + \dots + x_{i,10} \cdot \beta_{10}) = \exp(\beta_0 + \beta_1 + \lambda).$$

$$(\mathbf{y}_i, \mathbf{x}_{i0}) \Rightarrow \log\left(\frac{p_{i0,2}}{p_{i0,1}}\right) = \beta_0 + 0 \cdot \beta_1 + \dots + x_{i,10} \cdot \beta_{10} \Leftrightarrow$$

$$\frac{p_{i0,2}}{p_{i0,1}} = \exp(\beta_0 + \dots + x_{i,10} \cdot \beta_{10}) = \exp(\beta_0 + \lambda).$$

Portanto,

$$\frac{p_{i1,2}}{p_{i1,1}} = \exp(\beta_0 + \beta_1 + \lambda),$$

$$\frac{p_{i0,2}}{p_{i0,1}} = \exp(\beta_0 + \lambda).$$

Calculando a razão de chances, temos que:

$$\frac{\frac{p_{i1,2}}{p_{i1,1}}}{\frac{p_{i0,2}}{p_{i0,1}}} = \frac{\exp(\beta_0 + \beta_1 + \lambda)}{\exp(\beta_0 + \lambda)} = \exp(\beta_1)$$

Ou seja, $\exp(\beta_j)$ (exceto β_0) representa a razão de chances de ser Sambaqui quando se altera a variável $x_{i,j}$, *ceteris paribus*, em relação a quando não se altera, sendo as alterações:

- $x_{i,j}$ para uma categoria que não é a de referência ($x_{i,j}$ categórica);
- $x_{i,j} + 1$ (numérica).

Foi utilizado o AIC para a redução desse modelo e ele pôde ser reduzido a:

$$\log \left(\frac{p_{Sambaqui}}{p_{Itaipu}} \right) = 0,42 - 2,60. Coord_{Np} + 2,31. Coord_{Ep} + 0,86. Dist_p + 1,29.1(Elev > 30)$$

Esses são os mesmos coeficientes trazidos na Tabela 3 e as interpretações dos parâmetros estão na Seção 6.1.

Como mostrado anteriormente, conseguimos escrever apenas as probabilidades de cada filiação cultural:

$$p_{Sambaqui} = \frac{\exp(0,42 - 2,60. Y_p + 2,31. X_p + 0,86. Dist_p + 1,29.1(Elev > 30))}{1 + \exp(0,42 - 2,60. Y_p + 2,31. X_p + 0,86. Dist_p + 1,29.1(Elev > 30))'}$$

$$p_{Itaipu} = \frac{1}{1 + \exp(0,42 - 2,60. Y_p + 2,31. X_p + 0,86. Dist_p + 1,29.1(Elev > 30))}.$$

C.3. Modelo trinomial

Agora, vamos para o desenvolvimento do modelo trinomial. Analogamente ao modelo binomial, contamos com informações 135 sítios Cerâmicos ($n = 135$) com as filiações Tupiguarani ($p_{i,1}$), Aratu ($p_{i,2}$) e Una ($p_{i,3}$).

As variáveis utilizadas no modelo inicial (Cerâmico) foram dez, sendo as oito apresentadas Tabela C.1 mais as duas contidas na Tabela C.2:

Tabela C.2 Tabela das variáveis iniciais para a construção do modelo trinomial

Variável	Tipo	Parametrização
Componente ($k_{i,9}$)	Categórica (2)	$x_{i,11} = 1(k_{i,9} = "Comp_1")$ $x_{i,12} = 1(k_{i,10} = "TC_1")$,
Tipo cerâmico ($k_{i,10}$)	Categórica (4)	$x_{i,13} = 1(k_{i,10} = "TC_2")$, $x_{i,14} = 1(k_{i,10} = "TC_3")$

Para essa tabela, temos as mesmas considerações que as anteriores.

Sendo assim, o modelo Cerâmico pode ser escrito como:

$$\mathbf{y}_i | \mathbf{x}_i \sim \text{Multinomial}(1, \mathbf{p}_i), \quad i = 1, \dots, 135,$$

com,

$$\mathbf{y}_i = (y_{i,1}, y_{i,2}, y_{i,3}); \quad y_{i,1} + y_{i,2} + y_{i,3} = 1; \quad \mathbf{y}_i \text{ e } \mathbf{y}_j \text{ indep. para } i \neq j;$$

Sítio Tupiguarani: $\mathbf{y}_i = (1, 0, 0)$; Sítio Aratu: $\mathbf{y}_i = (0, 1, 0)$; Sítio Una: $\mathbf{y}_i = (0, 0, 1)$;

$$\mathbf{x}_i = (x_{i,1}, x_{i,2}, \dots, x_{i,14}); \quad \mathbf{p}_i = (p_{i,1}, p_{i,2}, p_{i,3}); \quad p_{i,1} + p_{i,2} + p_{i,3} = 1;$$

$p_{i,1}$: Probabilidade de ser Tupiguarani;

$p_{i,2}$: Probabilidade de ser Aratu; $p_{i,3}$: Probabilidade de ser Una.

Com a parametrização de logito de referência (Tupiguarani), temos:

$$\log\left(\frac{p_{i,2}}{p_{i,1}}\right) = \beta_0 + x_{i,1}\beta_1 + x_{i,2}\beta_2 + \dots + x_{i,14}\beta_{14}.$$

- Logaritmo da chance de ser Aratu em relação a ser Tupiguarani

$$\log\left(\frac{p_{i,3}}{p_{i,1}}\right) = \gamma_0 + x_{i,1}\gamma_1 + x_{i,2}\gamma_2 + \dots + x_{i,14}\gamma_{14}.$$

- Logaritmo da chance de ser Una em relação a ser Tupiguarani

Fazendo algumas transformações algébricas análogas às anteriores, temos que:

$$p_{i,2} = \exp(\beta_0 + x_{i,1}\beta_1 + x_{i,2}\beta_2 + \dots + x_{i,14}\beta_{14}) \cdot p_{i,1},$$

$$p_{i,3} = \exp(\gamma_0 + x_{i,1}\gamma_1 + x_{i,2}\gamma_2 + \dots + x_{i,14}\gamma_{14}) \cdot p_{i,1},$$

$$p_{i,1} = \frac{1}{1 + \exp(\beta_0 + x_{i,1}\beta_1 + \dots + x_{i,14}\beta_{14}) + \exp(\gamma_0 + x_{i,1}\gamma_1 + \dots + x_{i,14}\gamma_{14})}.$$

Além disso, os parâmetros β_j e γ_j do modelo têm interpretações análogas às anteriores.

O intercepto β_0 é o valor do logaritmo da chance de ser Aratu em relação a ser Tupiguarani para o sítio com as covariáveis numéricas iguais a média do conjunto de dados e as covariáveis categorizadas com o valor de referência.

Aplicando a exponencial no intercepto β_0 , temos o valor da chance:

$$\log\left(\frac{p_{i,2}}{p_{i,1}}\right) = \beta_0 + 0.\beta_1 + 0.\beta_2 + \dots + 0.\beta_{10} = \beta_0 \Leftrightarrow \frac{p_{i,2}}{p_{i,1}} = \exp(\beta_0).$$

O mesmo vale para o intercepto γ_0 , que é o valor para o logaritmo da chance de ser Una em relação a ser Tupiguarani.

Calculando a razão de chances como anteriormente, $\exp(\beta_j)$ (exceto β_0) representa a razão de chances de ser Aratu em relação a ser Tupiguarani quando se altera a variável $x_{i,j}$, *ceteris paribus*, em relação a quando não se altera, sendo as alterações:

- $x_{i,j}$ para uma categoria que não é a de referência ($x_{i,j}$ categórica);
- $x_{i,j} + 1$ (numérica).

O resultado é análogo para os parâmetros γ_j (exceto γ_0) para Una no lugar de Aratu.

Também utilizando o AIC para a seleção das variáveis que compõem o modelo, podemos reduzi-lo a:

$$\begin{aligned} \log\left(\frac{p_{Aratu}}{p_{Tupiguarani}}\right) &= -3,19 - 1,02 \cdot Coord_{Np} + 3,13 \cdot Coord_{Ep} - 0,51 \cdot Dist_p - \\ &\quad - 0,62 \cdot 1(Geo = "Planicies") + 3,28 \cdot 1(TC = "Simples e Dec. Plas.") \\ &\quad - 2,33 \cdot 1(TC = "Todos, Não se aplica") + 0,79 \cdot 1(TC = "Não espec.") \\ &\quad + 1,63 \cdot 1(MC = "Cerâmica e outro"). \end{aligned}$$

$$\begin{aligned}
\log \left(\frac{p_{Una}}{p_{Tupiguarani}} \right) = & -6,13 - 1,35 \cdot Coord_{Np} - 1,57 \cdot Coord_{Ep} - 1,59 \cdot Dist_p + \\
& + 4,14 \cdot 1(Geo = "Planicies") + 5,09 \cdot 1(TC = "Simples e Dec. Plas.") \\
& - 5,08 \cdot 1(TC = "Todos, Não se aplica") + 0,50 \cdot 1(TC = "Não espec.") \\
& - 0,08 \cdot 1(MC = "Cerâmica e outro").
\end{aligned}$$

Esses são os mesmos coeficientes trazidos na Tabela 6 e as interpretações dos parâmetros estão na Seção 6.2.

Análogo ao modelo binomial, conseguimos escrever apenas as probabilidades de cada filiação cultural:

$$\begin{aligned}
A = & -3,19 - 1,02 \cdot Y_p + 3,13 \cdot X_p - 0,51 \cdot Dist_p - 0,62 \cdot 1(Geo_1) \\
& + 3,28 \cdot 1(TC_1) - 2,33 \cdot 1(TC_2) + 0,79 \cdot 1(TC_3) + 1,63 \cdot 1(MP_1), \\
U = & -6,13 - 1,35 \cdot Y_p - 1,57 \cdot X_p - 1,59 \cdot Dist_p + 4,14 \cdot 1(Geo_1) \\
& + 5,09 \cdot 1(TC_1) - 0,88 \cdot 1(TC_2) + 0,50 \cdot 1(TC_3) - 0,08 \cdot 1(MP_1),
\end{aligned}$$

$$p_{Aratu} = \frac{\exp(A)}{1 + \exp(A) + \exp(U)},$$

$$p_{Una} = \frac{\exp(U)}{1 + \exp(A) + \exp(U)},$$

$$p_{Tupiguarani} = \frac{1}{1 + \exp(A) + \exp(U)}.$$

APÊNDICE D

Análise de resíduos e simulações

D.1. Análise de Resíduos para o modelo binomial

As análises de resíduos desenvolvidas para os modelos (ver Apêndice C) foram feitas via envelope simulado para a distribuição Multinomial, detecção de pontos de influência local e na locação/escala. O código implementado foi desenvolvido por Frederico Poleto baseado na teoria registrada por Gilberto Paula (IME-USP).

As Figuras D.1, D.2 e D.3 apresentam os resíduos com envelope simulado para 100 simulações com uma banda de confiança de 90% para o modelo binomial (para filiações não ceramistas). Em ambos os gráficos, os resíduos estão bem comportados dentro da banda de confiança, o que indica que o modelo adotado é adequado sob essa perspectiva.

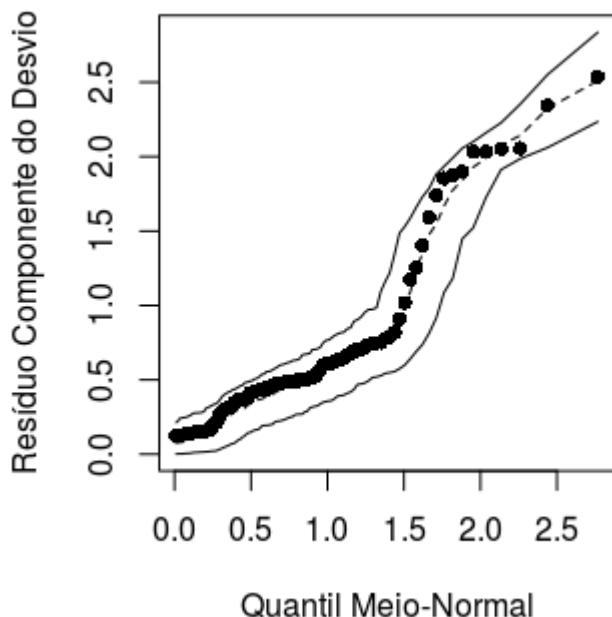


Figura D.1 Resíduo Componente do Desvio com banda de confiança de 90% para o modelo binomial

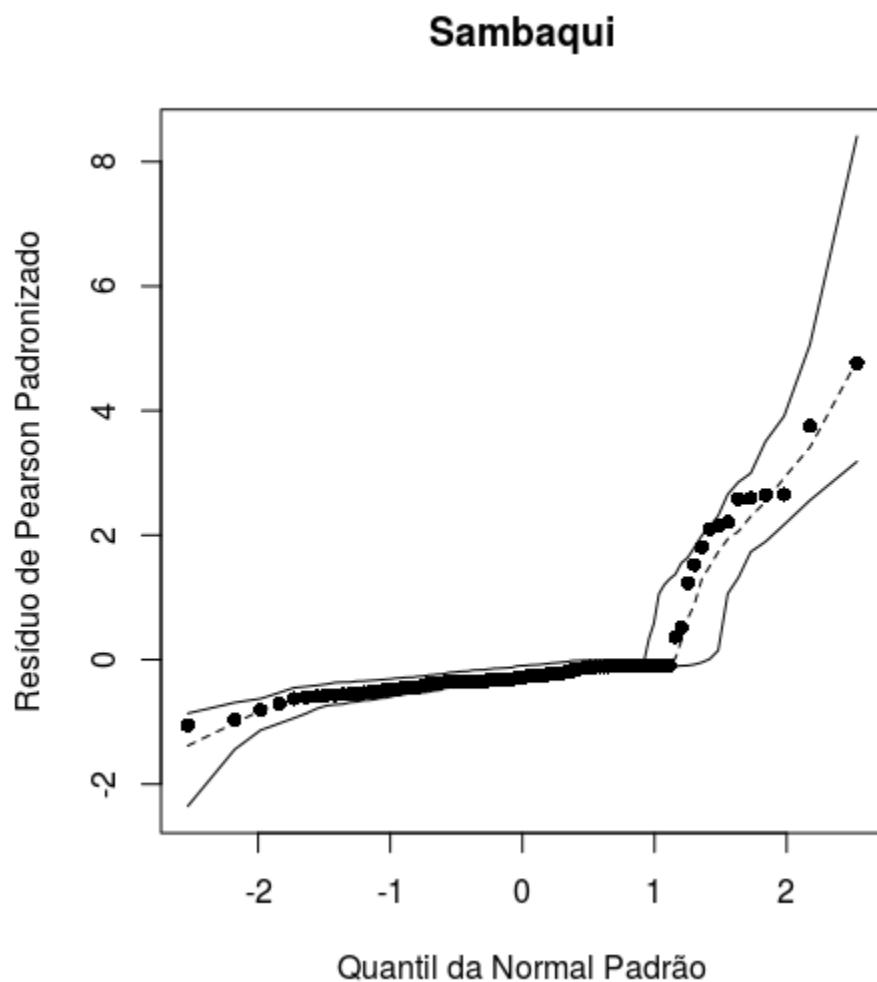


Figura D.2 Resíduo de Pearson Padronizado com banda de confiança de 90% para o modelo binomial

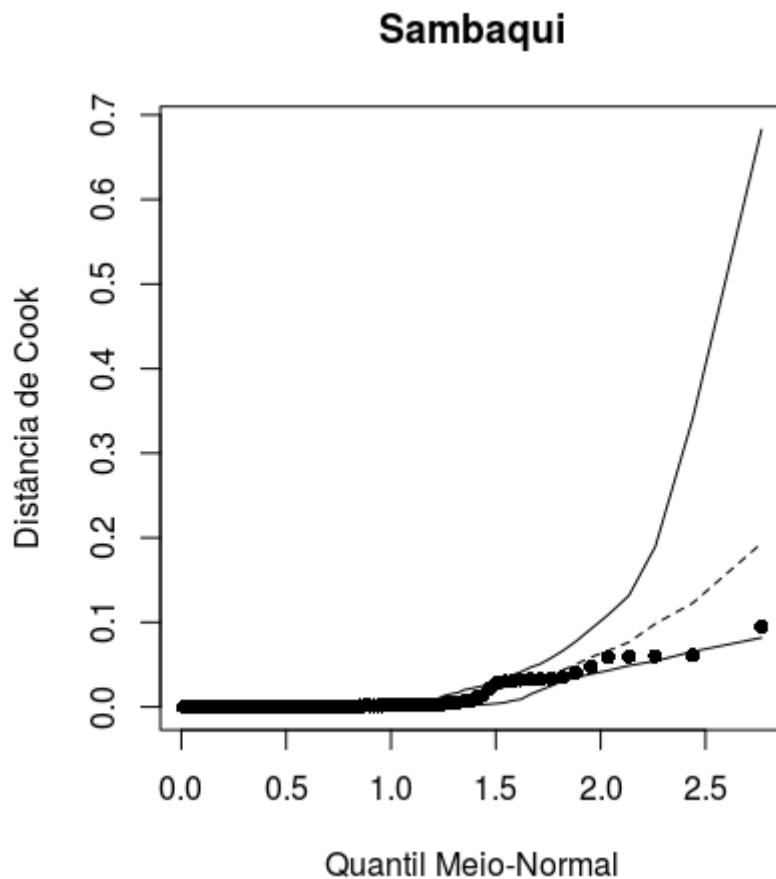


Figura D.3 Distância de Cook com banda de confiança de 90% para o modelo binomial

Já a Figura D.4 mostra dois gráficos de influência local, dois gráficos de influência na locação/escala e quatro gráficos para detecção de pontos de alavanca. Apesar de termos alguns pontos acima da linha tracejada, a escala dos gráficos nos leva a crer que as influências e discrepâncias são aceitáveis para o modelo binomial.

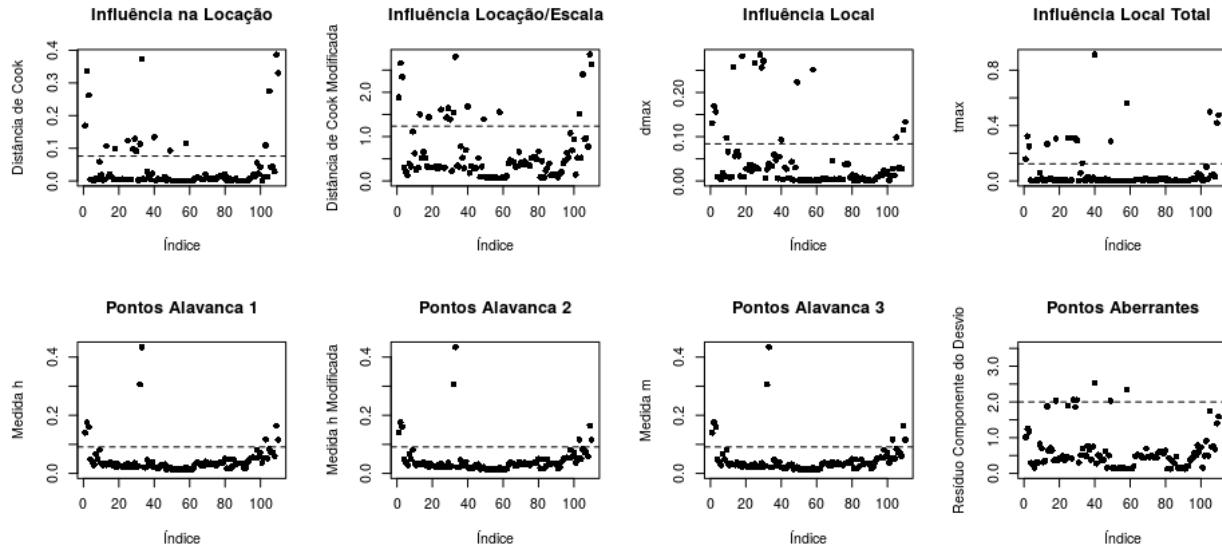


Figura D.4 Gráficos de influência local, global e detecção de pontos de alavancas para o modelo binomial

Assim sendo, concluímos que o modelo binomial, para as filiações não ceramistas está bem ajustado e pode ser considerado como um modelo apropriado.

D.2. Análise de Resíduos para o modelo trinomial

As Figuras D.5, D.6 e D.7 apresentam os resíduos com envelope simulado para 100 simulações com uma banda de confiança de 90% para o modelo trinomial (para filiações ceramistas). Os gráficos são feitos para Aratu e Una porque Tupiguarani é a filiação de referência da modelagem. Em ambos os gráficos, os resíduos estão bem comportados dentro da banda de confiança, o que indica que o modelo adotado é razoável sob essa perspectiva.

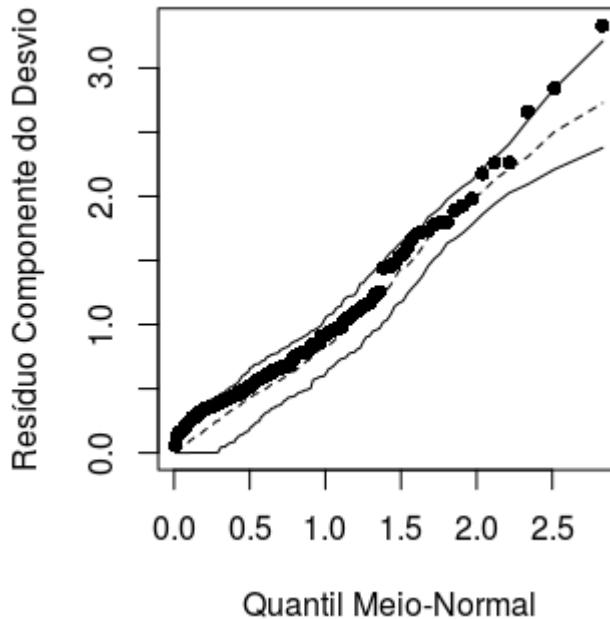


Figura D.5 Resíduo Componente do Desvio com banda de confiança de 90% para o modelo trinomial

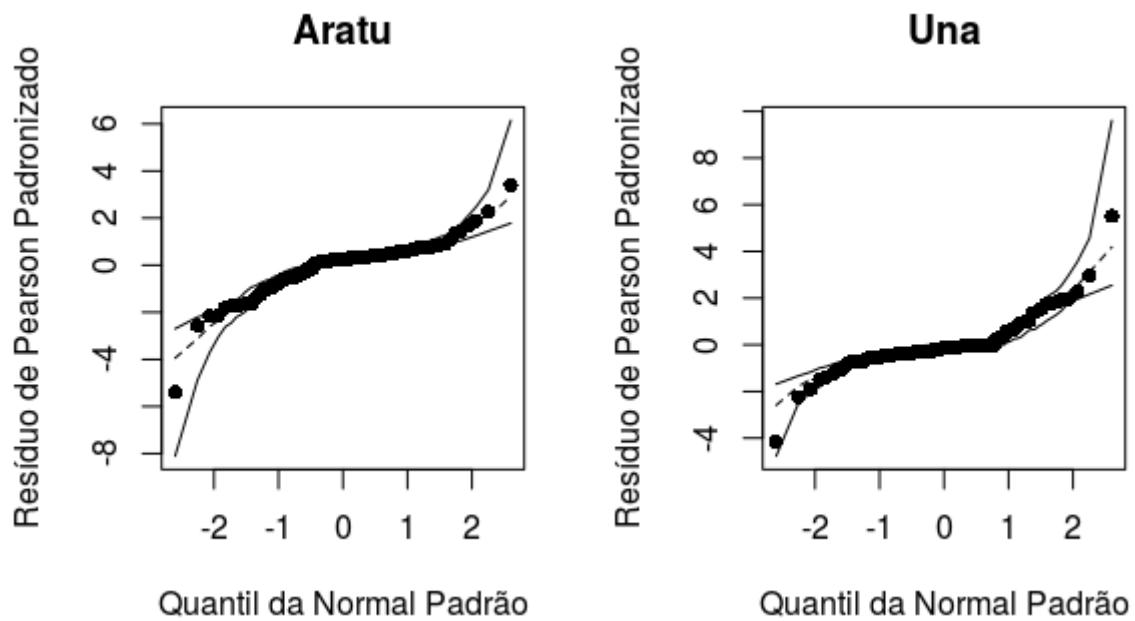


Figura D.6 Resíduo de Pearson Padronizado com banda de confiança de 90% para o modelo trinomial

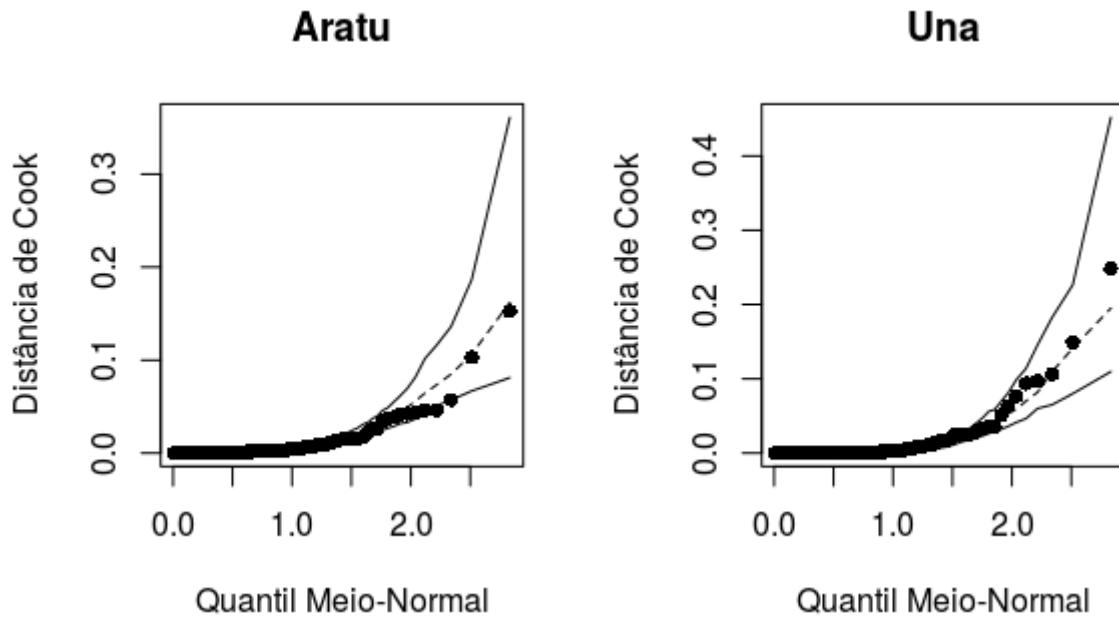


Figura D.7 Distância de Cook com banda de confiança de 90% para o modelo trinomial

Já a Figura D.8 mostra dois gráficos de influência local, dois gráficos de influência na locação/escala e quatro gráficos para detecção de pontos de alavanca. Apesar de alguns pontos estarem acima da linha tracejada, a escala dos gráficos nos leva a crer que as influências e discrepâncias são aceitáveis para o modelo trinomial.

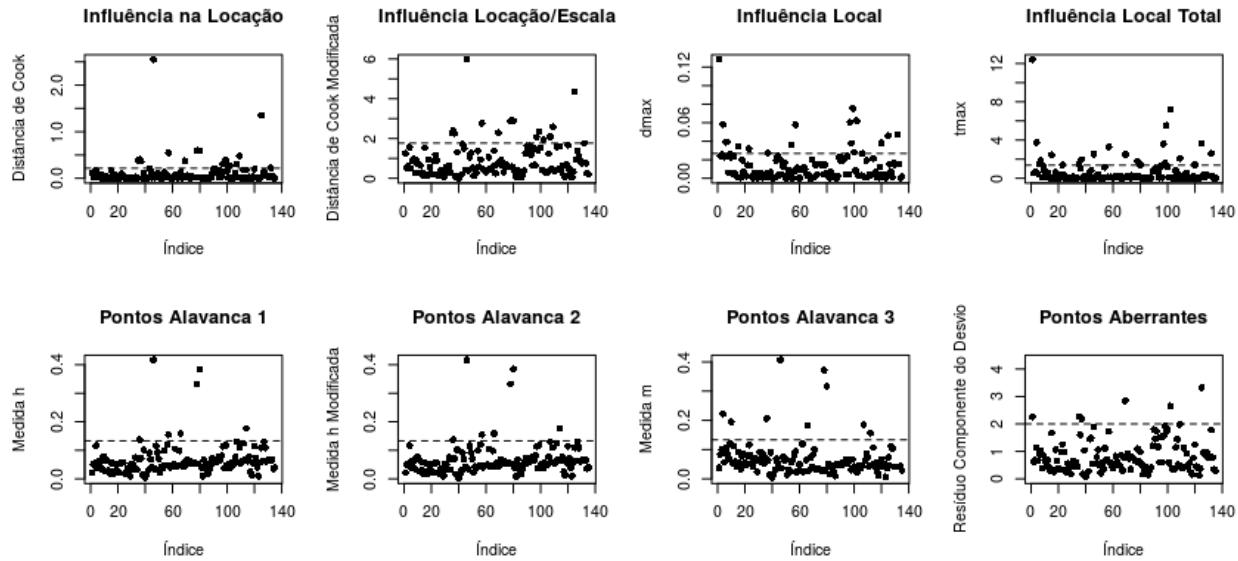


Figura D.8 Gráficos de influência local, global e detecção de pontos de alavanca para o modelo trinomial

Assim sendo, concluímos que o modelo trinomial, para as filiações ceramistas está bem ajustado e pode ser considerado como um modelo apropriado.

D.3. Validação do tamanho amostral

Além disso, via Simulação de Monte Carlo, validamos se o tamanho da amostra (n) é suficientemente grande para testar hipóteses com a Estatística de Wald, uma vez que, sob H_0 , se o valor-p convergir (em Distribuição) para uma distribuição Uniforme(0,1), então vale a aproximação assintótica para uma distribuição Qui-Quadrado (Sen, Singer & Pedroso de Lima, 2010).

Para tal, executamos o seguinte procedimento:

$$\mathbf{Y} = \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_n \end{bmatrix}, \text{ com } \mathbf{y}_i = (y_{i1}, \dots, y_{ik}) \quad \text{e} \quad \mathbf{X} = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_n \end{bmatrix}, \text{ com } \mathbf{x}_i = (x_{i1}, \dots, x_{ip})$$

- Ajustamos o modelo $[Y \sim X]$ e obtemos β_0 E.M.V. para β
- Assumindo β_0 como verdadeiro valor de β , obtemos para cada sítio a probabilidade de pertencer a cada filiação:

$$\mathbf{p}_o = \begin{bmatrix} \mathbf{p}_{01} \\ \mathbf{p}_{02} \\ \vdots \\ \mathbf{p}_{0n} \end{bmatrix} = \begin{bmatrix} f_1(\beta_0, x_1), f_2(\beta_0, x_1), \dots, f_k(\beta_0, x_1) \\ f_1(\beta_0, x_2), f_2(\beta_0, x_2), \dots, f_k(\beta_0, x_2) \\ \vdots \\ f_1(\beta_0, x_n), f_2(\beta_0, x_n), \dots, f_k(\beta_0, x_n) \end{bmatrix},$$

em que $f_j(\beta_0, x_i)$ são as funções com parâmetros β_0 e x_i para o cálculo de p_{0i} .

Assim sendo:

- 1) Geramos uma amostra aleatória Y_{sim} utilizando \mathbf{p}_0 (garantindo que tenhamos pelo menos uma observação por casela);
- 2) Ajustamos o modelo $[Y_{sim} \sim X]$ obtendo o $\hat{\beta}_{sim}$ E.M.V. de β ;
- 3) Fazemos o teste $H_0: C\beta = C\beta_0$ e obtemos W (Estatística de Wald) com o valor-p, com C matriz de posto r

- Repetimos 1) a 3) 10.000 vezes e plotamos o histograma do valor-p.
- Então, sob H_0 , $W \xrightarrow{D} \chi^2_{(r)} \Leftrightarrow$ Valor-p $\xrightarrow{D} U(0,1)$

Com esse procedimento, validamos os tamanhos amostrais com histogramas de valores-p. As barras dos histogramas estavam bem semelhantes à distribuição Uniforme(0,1), indicando que o tamanho amostral é suficiente para se verificar H_0 .

As Figuras D.9 e D.10 são exemplos de simulações que fizemos para os interceptos do modelo trinomial. Nesse caso, fizemos o teste conjunto em que a hipótese nula é $H_0: \beta = \beta_0$. Ambos se assemelham a uma distribuição Uniforme(0,1), bem como os outros que não foram anexados nesse relatório,

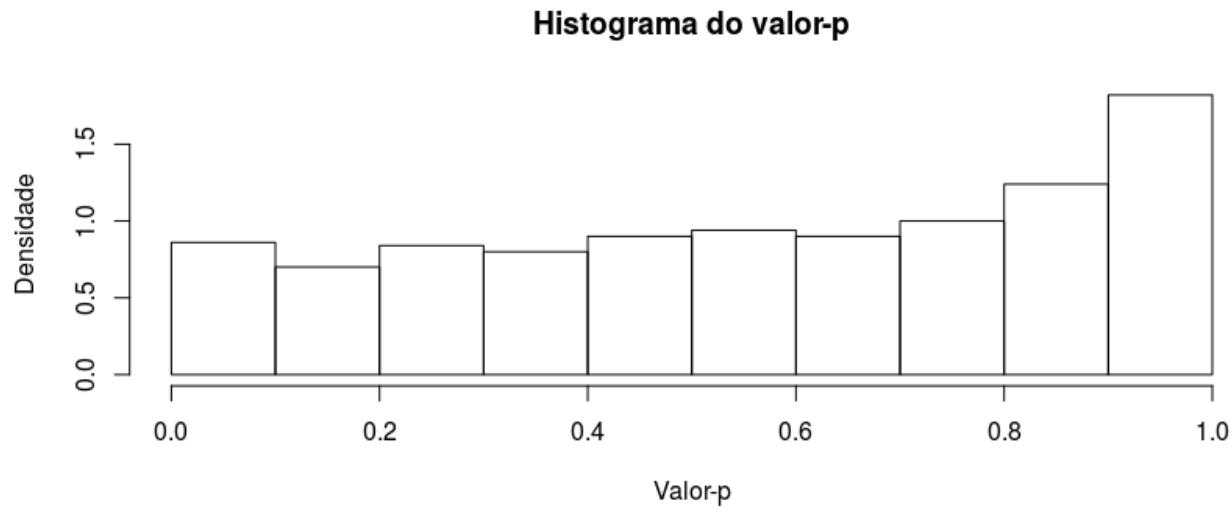


Figura D.9 Histograma do valor-p para as simulações do teste conjunto para o modelo trinomial

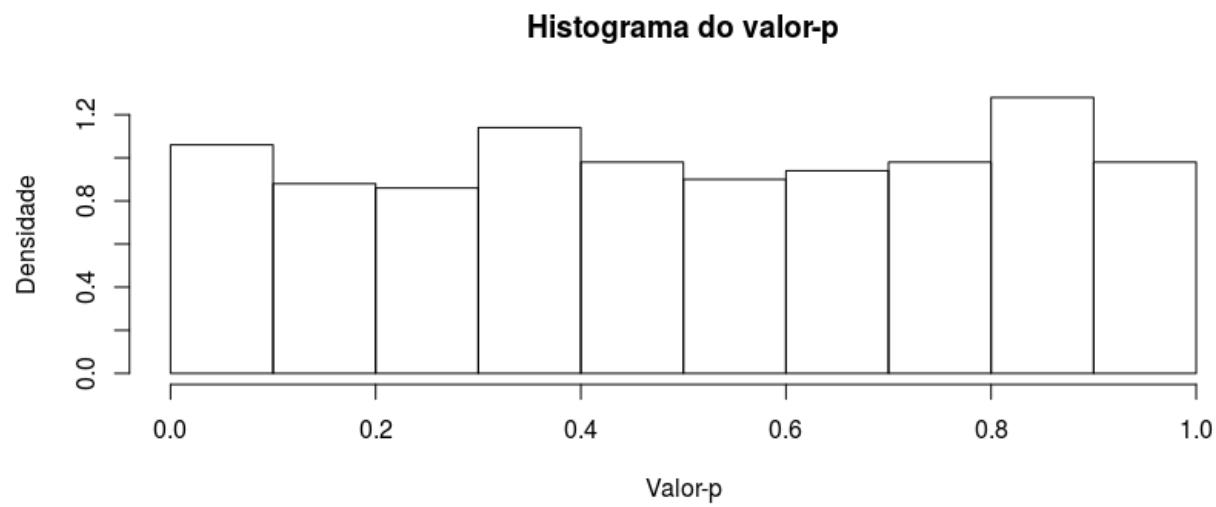


Figura D.10 Histograma do valor-p para as simulações do teste conjunto para o modelo binomial

ANEXO

Imagens de diferentes tipos de Material cultural associados a cada Filiação cultural. Todas as imagens foram disponibilizadas pela pesquisadora

Sambaqui



Figura AN.1 Material lítico: Quebra coquinho



Figura AN.2 Material lítico: Pré-ponta de quartzo



Figura AN.3 Exemplo: Formato de Sambaqui

Itaipu



Figura AN.4 À esquerda: Materiais líticos. À direita: Dentes de peixe e artefatos ósseos



Figura AN.5 Material lítico: Pré-ponta de quartzo



Figura AN.6 Exemplo: Sítio arqueológico

Aratu

Figura AN.7 Cerâmica: Cacos de cerâmica pré-colonial Aratu com decoração plástica

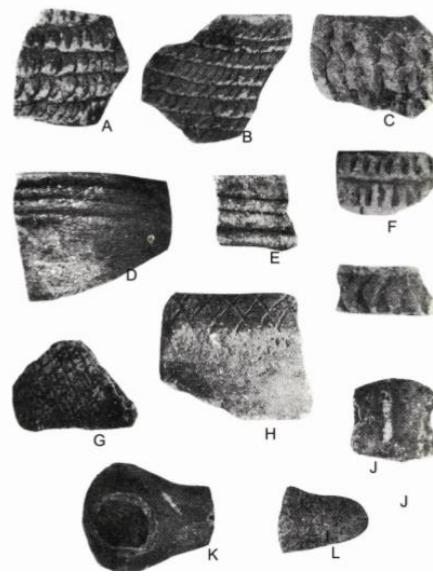


Figura AN.8 Cerâmica: Cacos de cerâmica pré-colonial Aratu com decoração plástica e simples

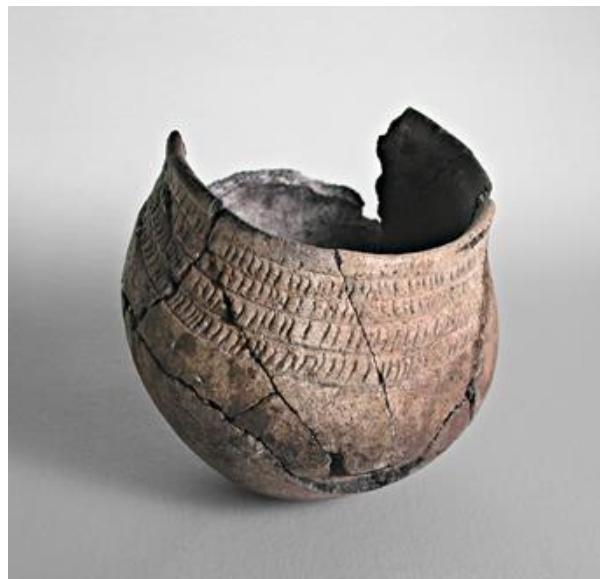


Figura AN.9 Cerâmica: Vaso de cerâmica pré-colonial Aratu com decoração plástica

Tupiguarani



Figura AN.10 Cerâmica: Caco de Cerâmica pré-colonial com decoração pintada

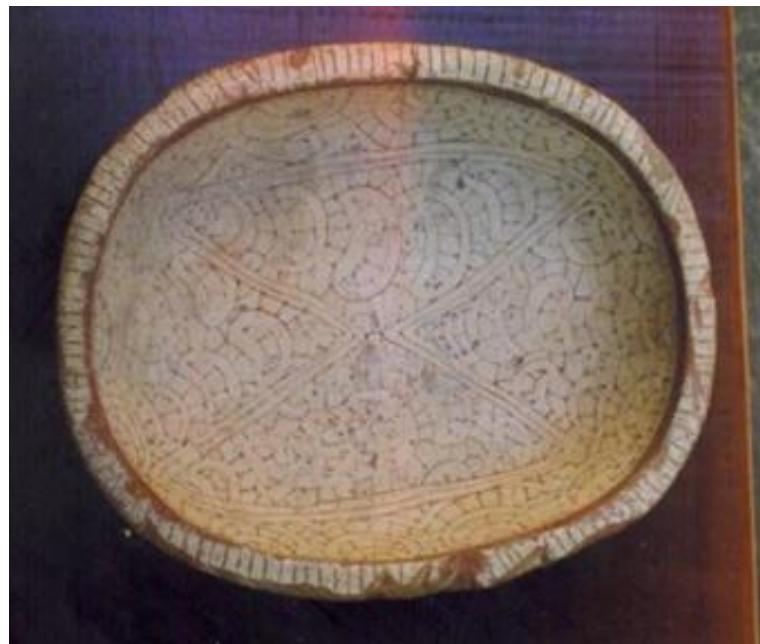


Figura AN.11 Cerâmica: Vasilhame de Cerâmica pré-colonial com decoração pintada



Figura AN.12 Cerâmica: Urna funerária

Una



Figura AN.13 Cerâmica: Cacos de cerâmica pré-colonial com decoração plástica e simples



Figura AN.14 Cerâmica: Fragmento de vaso globular com decoração simples