



Universidade Federal de São Carlos  
Centro de Educação e Ciências Humanas  
Departamento de Letras



Pedro Sabino

**ESTATÍSTICAS DE FREQUÊNCIA FUNDAMENTAL DA POPULAÇÃO MASCULINA  
DA CIDADE DE SÃO PAULO PARA USO EM FONÉTICA FORENSE**

São Carlos - SP

2025

Pedro Sabino

ESTATÍSTICAS DE FREQUÊNCIA FUNDAMENTAL DA POPULAÇÃO MASCULINA DA  
CIDADE DE SÃO PAULO PARA USO EM FONÉTICA FORENSE

Trabalho de Conclusão de Curso apresentado  
ao Curso de Bacharelado em Linguística da  
Universidade Federal de São Carlos.

orientador: Prof. Dr. Pablo Arantes

coorientadora: Prof<sup>a</sup> Dr<sup>a</sup>. Renata Regina Passetti

São Carlos - SP

2025

Sabino, Pedro.

S116 Estatísticas de frequência fundamental da população masculina da cidade de São Paulo para uso em fonética forense / Pedro Sabino. — 2025.

48 f.

Trabalho de Conclusão de Curso (Bacharelado em Linguística) — Universidade Federal de São Carlos, São Carlos, 2025.

1. Fonética forense. 2. Frequência fundamental. 3. Razão de verossimilhança. 4. Fala paulista. I. Título.

CDD 410.

UNIVERSIDADE FEDERAL DE SÃO CARLOS  
DEPARTAMENTO DE LETRAS  
BACHARELADO EM LINGUÍSTICA

**FOLHA DE APROVAÇÃO**

Pedro Sabino

Estatísticas de frequência fundamental da população masculina da cidade de São Paulo para uso em fonética forense

Trabalho de conclusão de curso apresentado como requisito parcial para a obtenção do título de Bacharel em Linguística pela Universidade Federal de São Carlos.

Aprovado em: 12/12/2025.

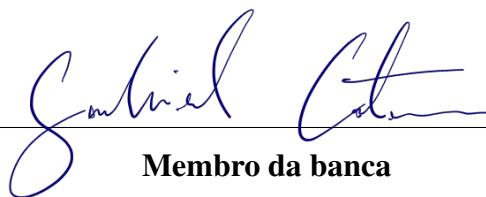
**BANCA AVALIADORA**

---

**Orientador**

Prof. Dr. Pablo Arantes  
Universidade Federal de São Carlos

---



**Membro da banca**

Prof. Gabriel Catani  
Universidade Federal de São Carlos

# AGRADECIMENTOS

A realização deste trabalho só foi possível graças ao apoio e à presença de muitas pessoas que, de diferentes formas, contribuíram para que esta trajetória fosse possível.

Primeiramente agradeço imensamente aos meus pais, que com muito esforço, paciência e amor incondicional, me deram o privilégio de dedicar-me integralmente à graduação, mesmo estando em outro estado. O apoio que vocês me deram foi fundamental para que eu pudesse seguir com tranquilidade e foco, mesmo diante dos desafios do dia a dia e da imensa saudade de casa.

Da mesma forma, agradeço aos meus amigos, que se tornaram uma segunda família em São Carlos. A convivência, o companheirismo e as conversas foram essenciais para tornar o percurso da graduação bem mais leve. A presença de vocês nos momentos mais desafiadores trouxe força, risadas e acolhimento.

Expresso também minha sincera gratidão ao meu orientador, Pablo Arantes, e à minha coorientadora, Renata Regina Passetti, pelo cuidado, paciência e dedicação ao longo de todas as etapas deste trabalho. A colaboração, o conhecimento compartilhado e as contribuições de ambos foram fundamentais para o amadurecimento deste trabalho e para o meu crescimento acadêmico e pessoal.

Por fim, agradeço a todos que, de alguma forma, estiveram presentes durante essa jornada — colegas, professores e demais pessoas que contribuíram direta ou indiretamente para a conclusão deste trabalho.

## RESUMO

Este trabalho teve como objetivo principal gerar estatísticas de referência da frequência fundamental ( $f_0$ ) da fala da população masculina da cidade de São Paulo-SP como recurso a ser aplicado na prática da fonética forense. As amostras utilizadas neste projeto de pesquisa foram retiradas do acervo do Projeto SP2010, que foi desenvolvido pelo Grupo de Estudos e Pesquisa em Sociolinguística da Universidade de São Paulo (USP). As amostras foram estratificadas por fatores sociodemográficos como escolaridade, zona geográfica e faixa etária.

Durante a condução deste projeto, foram selecionados trechos de áudio com duração em torno de 45 segundos de cada amostra de fala dos quais foram extraídos os contornos de  $f_0$ . A partir dos contornos foram calculados descritores estatísticos de  $f_0$  agregados pelos fatores sociodemográficos escolaridade, zona geográfica e faixa etária dos falantes.

Os sumários estatísticos resultantes são o principal resultado obtido pelo trabalho e podem ser usados em exames de comparação de locutor que usem como metodologia para expressão de sua conclusão a metodologia baseada na razão de verossimilhança (*likelihood ratio*). Além dos sumários estatísticos analisamos a influência dos fatores sociodemográficos de estratificação definidos pelo Projeto SP2010 sobre alguns dos descritores estatísticos. Os resultados dessa análise indicaram significância estatística entre as variáveis faixas etária e zonas geográfica, com a hipótese de variação sociolinguística entre os falantes das zonas geográficas e valores de  $f_0$  mais baixos entre o grupo de falantes com mais de 60 anos.

**Palavras-chave:** Fonética forense; Frequência fundamental; Razão de verossimilhança; Fala paulista.

## ABSTRACT

This study aimed to establish reference statistics for the fundamental frequency ( $f_0$ ) of male speech in the city of São Paulo, Brazil, with a focus on applications in forensic phonetics. The speech samples used in this research were obtained from the SP2010 Project database, developed by the Sociolinguistics Research Group at the University of São Paulo (USP). These samples were stratified according to sociodemographic factors such as education level, geographic zone, and age group.

Throughout the course of this project, 45-second segments were selected from each speech sample, from which  $f_0$  contours were extracted. These contours were then used to calculate statistical descriptors of  $f_0$ , aggregated according to the sociodemographic factors of education level, geographic zone, and age group of the speakers.

The resulting statistical summaries constitute the main outcome of this study and can serve as reference data in speaker comparison examinations that apply the likelihood ratio methodology. Beyond generating these summaries, the study also examined the influence of the sociodemographic stratification factors established by the SP2010 Project on selected statistical descriptors. The analysis revealed statistically significant relationships between age group and geographic zone, suggesting potential sociophonetic variation among speakers from different regions, as well as lower  $f_0$  values among those aged over 60.

**Keywords:** Forensic phonetics; Fundamental frequency; Likelihood ratio; São Paulo.

## LISTA DE FIGURAS

|            |   |    |
|------------|---|----|
| Figura 1 – | Gráfico de simulação do cálculo da razão de verossimilhança. . . . .  | 16 |
| Figura 2 – | Gráfico de simulação do cálculo da razão de verossimilhança. . . . .  | 17 |
| Figura 3 – | Gráfico de simulação do cálculo da razão de verossimilhança. . . . .  | 18 |
| Figura 4 – | Diagramas de caixa dos descritores de tendência central $f_0$ em função da escala de mensuração e do nível de escolaridade dos participantes. . | 26 |
| Figura 5 – | Diagramas de caixa dos descritores de tendência central $f_0$ em função da escala de mensuração e dos estilos de elocução dos participantes. .  | 29 |
| Figura 6 – | Diagramas de caixa dos descritores de tendência central $f_0$ em função da escala de mensuração e das zonas geográficas da cidade de São Paulo. | 32 |
| Figura 7 – | Diagramas de caixa dos descritores de tendência central $f_0$ em função da escala de mensuração e das faixas etárias dos participantes. . . . . | 35 |

## LISTA DE TABELAS

|            |  |    |
|------------|--|----|
| Tabela 1 – | Medidas de $f_0$ em Hertz (Hz) e Semitons (ST), por nível de Escolaridade. | 25 |
| Tabela 2 – | Medidas de $f_0$ em Hertz (Hz) e Semitons (ST) por Estilo de Elocução.     | 28 |
| Tabela 3 – | Medidas de $f_0$ em Hertz (Hz) e Semitons (ST) por Zona Geográfica. . .    | 31 |
| Tabela 4 – | Medidas de $f_0$ em Hertz (Hz) e Semitons (ST) por Faixa Etária. . . . .   | 34 |

# Sumário

|          |   |           |
|----------|---|-----------|
| <b>1</b> | <b>INTRODUÇÃO</b>   | <b>9</b>  |
| 1.1      | O Que é Fonética Forense? . . . . .   | 9         |
| 1.2      | Sobre o Exame de Comparação de Locutores . . . . .  | 9         |
| 1.3      | Parâmetro acústico de f0 na fonética forense . . . . .  | 10        |
| 1.4      | Definição de população relevante . . . . .  | 10        |
| 1.5      | Entendendo a importância da similaridade e da tipicidade na comparação de locutor . . . . .     | 11        |
| 1.6      | Importância da dialetologia na fonética . . . . .   | 12        |
| <b>2</b> | <b>FUNDAMENTAÇÃO TEÓRICA</b>  | <b>14</b> |
| 2.1      | Arcabouço da razão de verossimilhança . . . . .   | 14        |
| 2.2      | Aplicação da LR . . . . .   | 15        |
| <b>3</b> | <b>METODOLOGIA</b>  | <b>19</b> |
| 3.1      | Sobre o corpus utilizado . . . . .  | 19        |
| 3.2      | Seleção das amostras para análise . . . . .   | 19        |
| 3.3      | Variáveis analisadas e métodos de extração . . . . .  | 20        |
| 3.4      | Análise dos dados . . . . .   | 21        |
| 3.5      | Análise estatística dos valores de f0 . . . . .   | 22        |
| 3.6      | Testes estatísticos . . . . .   | 23        |
| <b>4</b> | <b>RESULTADOS E DISCUSSÃO</b>   | <b>25</b> |
| 4.1      | Efeito da escolaridade sobre os descritores estatísticos . . . . .                              | 25        |
| 4.2      | Efeito dos estilos de elocução sobre os descritores estatísticos . . . . .                      | 28        |
| 4.3      | Efeito das zonas geográficas da cidade de São Paulo sobre os descritores estatísticos . . . . . | 30        |
| 4.4      | Efeito de diferentes faixas etárias sobre os descritores estatísticos . . . . .                 | 34        |
| <b>5</b> | <b>CONSIDERAÇÕES FINAIS</b>   | <b>38</b> |
|          | <b>ANEXOS</b>   | <b>42</b> |
| <b>A</b> | <b>CÓDIGO SCRIPT SLICE MERGE</b>  | <b>42</b> |
| <b>B</b> | <b>CÓDIGO TABELA DE SIMULAÇÕES</b>  | <b>44</b> |
| <b>C</b> | <b>CÓDIGO SCRIPT PITCH TO TABLE</b>   | <b>45</b> |

# 1 INTRODUÇÃO

## 1.1 O Que é Fonética Forense?

A fonética forense constitui uma área da linguística forense, cuja base metodológica está relacionada à análises fonéticas e acústicas. Essa área do conhecimento dedica-se à análise técnica de amostras de áudio coletadas em situações sob ordens judiciais, contribuindo, em conjunto com a perícia, para a elucidação de casos em que o infrator foi apenas ouvido, considerando-se que o indivíduo submetido à análise fonético-forense pode não ser o responsável pela infração investigada.

Uma das tarefas da área consiste na realização de análises dos sinais acústicos da fala a fim de auxiliar na comparação e identificação de suspeitos. No entanto, as aplicações da fonética forense vão além da comparação de locutores. As tarefas nessa área incluem a elaboração de perfis linguísticos e fonéticos, a caracterização de variedades regionais ou sociolinguísticas, a avaliação de condições técnicas de gravação e a investigação de fatores fisiológicos que afetam a fala.

No âmbito de suas aplicações, a fonética forense realiza a comparação de locutores, que consiste na análise técnica de duas ou mais amostras de voz para determinar a probabilidade de terem sido produzidas pelo mesmo indivíduo. Para isso, a frequência fundamental ( $f_0$ ) se destaca como um dos parâmetros cruciais para a caracterização individual da voz. (HUGHES, 2014).

## 1.2 Sobre o Exame de Comparação de Locutores

Entre as principais atividades da fonética forense está a comparação de locutores (JESSEN, 2008), referida a partir daqui por CL, que envolve a análise detalhada de diferentes características vocais e acústicas, com a finalidade de avaliar a probabilidade de que duas ou mais amostras de fala tenham sido produzidas pelo mesmo falante. Nesse procedimento, uma amostra de voz cujo locutor tem identidade desconhecida (chamada de amostra *questionada*) é confrontada com amostras de *referência*, fornecidas por locutores considerados suspeitos de serem o autor da voz de identidade desconhecida em um contexto de investigação judicial.

Nessas amostras selecionadas, são coletados e analisados parâmetros fonéticos segmentais, suprasegmentais, além de parâmetros associados a outros níveis de análise linguística. O objetivo é determinar em que grau as amostras questionadas e as de referência podem ser

atribuídas à mesma fonte produtora ou a fontes distintas (ou seja, se o locutor da amostra suspeita é também a fonte da amostra questionada ou se as amostras questionada e suspeita não foram produzidas pelo mesmo locutor). Normalmente, a amostra questionada é obtida por meio de interceptações telefônicas, escuta ambiental e, mais recentemente, áudios gravados em aplicativos de mensagem, como *WhatsApp* e *Telegram* (PASSETTI; MADUREIRA; BARBOSA, 2022), enquanto a amostra de referência é coletada por peritos, em condições controladas, com equipamentos de alta qualidade de indivíduos cuja identidade já é conhecida (i.e. de suspeitos) (NOLAN, 1983).

No Brasil, a prática do exame de comparação de locutores geralmente se inicia através de solicitação do Ministério Público ou de solicitação da defesa, em que o exame é ordenado pelo juiz. Após isso, a condução do exame é feita por peritos que realizam uma análise auditiva e fonético-acústica combinadas. Na análise auditiva, através das amostras disponíveis, são analisados parâmetros segmentais e suprasegmentais. Na análise fonético-acústica, o perito analisa parâmetros físicos do sinal de fala usando softwares especializados (GOLD; FRENCH, 2011).

### **1.3 Parâmetro acústico de f0 na fonética forense**

Entre os parâmetros suprasegmentais mais relevantes em exames de CL está a frequência fundamental (f0) (GOLD; FRENCH, 2011), que é o correspondente acústico da frequência de vibração das pregas vocais e, do ponto de vista perceptivo, é um dos fatores responsáveis por gerar a sensação de tonalidade na fala (*pitch* em inglês), em que, f0 mais alto, em geral equivale a sensação de fala mais aguda, enquanto f0 mais baixo, sensação de fala mais grave.

A relevância de f0 para a CL em contexto forense se justifica pela evidência de que características relevantes das pregas vocais, como seu comprimento, peso e espessura, afetam a produção da fala, sendo próprias de cada indivíduo. Além disso, trata-se de uma ferramenta capaz de ajudar a caracterizar o falante em contextos legais e auxiliar, juntamente com outros dados, informações sobre ele, como sua origem regional, idade e sexo.

Essa relação entre a anatomia e a fisiologia com a f0, no entanto, não é direta, uma vez que os padrões de f0 observados na fala também sofrem influência de fatores paralinguísticos, como estado emocional do locutor, além de fatores linguísticos estruturais (NOLAN, 1983) que também devem ser levados em conta. Tendo isso em vista, está justificada a escolha da frequência fundamental (f0) como foco do projeto.

### **1.4 Definição de população relevante**

Para que se consiga aplicar a razão de verossimilhança, a definição de uma “população relevante” é fundamental, pois é a partir dessa população que a tipicidade das características de

fala é avaliada, permitindo aos peritos contextualizar a prevalência de um traço específico na fala do suspeito. Segundo a literatura, essa população pode ser entendida de duas maneiras.

A primeira, sugerida por Morrison (2012), consiste em falantes que possuem a voz suficientemente parecida com a voz do suspeito, de forma que pessoas leigas, isto é, não especialistas, consigam perceber essa similaridade entre as vozes das gravações questionada e suspeita. Somente após isso, um especialista analisaria as amostras, atestando se as vozes apontam ou não para a mesma origem. Desse ponto de vista, a população relevante seria concebida como o conjunto de indivíduos cujas vozes soam suficientemente parecidas com a voz do suspeito, de modo que um leigo incluiria estas amostras no exame de comparação de locutor.

A segunda concepção, sugerida por Hughes (2014), trata-se de um conceito de relevância lógica, em que a população relevante deve ser entendida como o conjunto de pessoas que tenham o mesmo sexo e sejam falantes do mesmo idioma do locutor que produziu a amostra questionada, informações essas que devem ser inferidas a partir da escuta das amostras analisadas. Pode-se definir a população de forma mais estreita levando em conta variáveis como proveniência regional, idade, escolaridade, renda, entre outras conforme seja possível justificar a adoção dessas estratificações a partir da análise de traços linguísticos e/ou fonéticos presentes nas amostras em comparação.

Segundo Morrison (2010), apenas um grau alto de similaridade entre as amostras não é suficiente para corroborar a hipótese de mesma origem em uma CL, pois pode ser que a semelhança encontrada seja muito comum entre os indivíduos de certa população relevante. Sendo assim, se houver a conjunção entre um grau alto de semelhança e baixa tipicidade dessas semelhanças na população, isso fortalece a hipótese de a amostra questionada ter a mesma origem da amostra de referência. Por outro lado, quanto maior for a tipicidade e menor a similaridade, maior é a probabilidade de as amostras serem de origem diferente.

Desta forma, é fundamental conhecer a distribuição estatística da  $f_0$ , bem como de parâmetros linguístico-fonéticos comumente avaliados em exames de CL, considerando populações de referência para a aplicação do método da razão de verossimilhança.

## **1.5 Entendendo a importância da similaridade e da tipicidade na comparação de locutor**

No exame de CL, avaliar apenas a semelhança entre a amostra questionada e a amostra de referência não é suficiente, uma vez que essa abordagem pode introduzir vieses na interpretação dos resultados. Para que seja possível encontrar resultados robustos, é necessário considerar simultaneamente não só a similaridade entre as amostras mas também a tipicidade dos traços observados, ou seja, quanto esses traços são frequentes. Portanto, a tipicidade remete diretamente à necessidade de definir e conhecer as características da população relevante à qual pertencem tanto o suspeito quanto o locutor da gravação questionada.

Essa comparação entre similaridade e tipicidade é feita através da razão de verossimilhança (*likelihood ratio*, em inglês, LR abreviadamente), que avalia o peso da evidência a partir da comparação entre as amostras de *referência* e amostra *suspeita*. Para que a LR seja aplicada de forma adequada, é fundamental conhecer a distribuição estatística de parâmetros linguístico-fonéticos comumente avaliados em exames de comparação de locutores, como a f0, considerando populações de referência para a aplicação do método da razão de verossimilhança (HUGHES, 2014).

O mapeamento dessa distribuição estatística é importante para reduzir possíveis vieses na análise do exame de CL, que poderiam favorecer a hipótese de que as amostras de fala comparadas têm a mesma origem. Nesse ponto, o grau de tipicidade desempenha papel central na avaliação justa e precisa do valor probatório de evidências de fala, uma vez que indica o quão típicos ou atípicos são os traços examinados no teste de CL em relação à população relevante, influenciando diretamente o cálculo da razão de verossimilhança. Além disso, a consideração da tipicidade das características analisadas é indispensável para a interpretação das medidas de f0 dentro de um determinado grupo populacional, contribuindo para a correta delimitação do conjunto de possíveis suspeitos.

Se o grau de tipicidade não for levado em conta, a avaliação do valor probatório da evidência pode ser distorcida, já que características comuns podem gerar a falsa impressão de correspondência entre as amostras. A partir disso, a LR pode ser aplicada, e, através dela é possível quantificar a força da evidência ao comparar a probabilidade de as diferenças observadas entre as amostras ocorrerem sob a hipótese de mesma origem ou de origem diferente.

Nesse sentido, a utilização de estatísticas de distribuição de parâmetros fonético-acústicos, retiradas de referências populacionais, torna-se essencial para a comparação de locutores, especialmente em abordagens baseadas na razão de verossimilhança.

## 1.6 Importância da dialetologia na fonética

A Dialetologia é uma área da Linguística que investiga as variações linguísticas, com ênfase na distribuição geográfica dessas variações. Essa linha teórica possui duas ênfases de análise, a primeira é descrever a distribuição da língua em um determinado espaço geográfico e a segunda é analisar de qual maneira se aplicam os fatores sociais na presença de variantes dessa mesma língua (CARDOSO, 2016).

Trazendo para o cenário da fonética forense, a Dialetologia serve de contribuição no contexto de mapeamento e definição da população relevante, estabelecendo critérios sociolinguísticos, como, por exemplo, a tipicidade da variante da língua que é falada em certa região.

Desta forma, as duas áreas atuando em conjunto podem auxiliar na produção de perfis linguísticos detalhados, que podem servir de ferramenta em investigações criminais, de maneira a ajudar a inferir aspectos da identidade e da origem do falante, desta forma colaborando para a boa condução de exames forenses. Além disso, este trabalho contribui para a Dialetologia ao

fornecer uma caracterização estatística da distribuição de um parâmetro acústico relevante para a variação linguística.

A escolha da cidade de São Paulo como foco deste estudo justifica-se por ser a cidade mais populosa do país, com uma população de aproximadamente 11,5 milhões de habitantes (IBGE, 2022), desta forma, a cidade constitui uma população relevante do ponto de vista criminalístico. Com uma população tão grande, as chances de crimes envolvendo pessoas dessa cidade aumenta e, com isso, a importância de conhecer as estatísticas dessa população.

## 2 FUNDAMENTAÇÃO TEÓRICA

### 2.1 Arcabouço da razão de verossimilhança

No exame de CL, após a análise das amostras questionada e de referência, procede-se à avaliação das semelhanças e diferenças observadas entre os parâmetros fonéticos examinados. Essa etapa tem como finalidade determinar se essas características configuram a evidência como favorável ou desfavorável à hipótese de mesma origem, isto é, se o suspeito é o autor de ambas as amostras, ou se as amostras possuem origens diferentes, de maneira em que o suspeito não seria o autor da amostra questionada.

Por conta disso, torna-se necessário estabelecer o peso probatório da evidência, a literatura apresenta uma tendência maior na utilização do arcabouço bayesiano, sendo apresentado também como *likelihood-ratio framework*, por se tratar de uma alternativa que supera as limitações de outros arcabouços de decisão, sendo denominada como mudança de paradigma (MORRISON, 2010).

A principal característica do arcabouço da razão de verossimilhança está na avaliação da força probatória da evidência pericial por meio da comparação simultânea entre duas probabilidades condicionais associadas à evidência analisada. Essa evidência corresponde às mensurações das diferenças e semelhanças entre as amostras de origem questionada e de referência.

Durante a aplicação desta tarefa, o perito avalia a força probatória das evidências extraídas dos dados vocais e compara as probabilidades de duas hipóteses distintas. A primeira hipótese sustenta que as amostras de referência e questionada estão ligadas à mesma fonte, ou seja, foram produzidas pelo mesmo autor. A segunda hipótese, por sua vez, supõe que as amostras têm origens diferentes e não foram produzidas pela mesma pessoa.

É importante enfatizar que, a responsabilidade de considerar a culpabilidade do suspeito pertence aos jurados e juízes, que consideram além do laudo técnico gerado pelo perito, outras evidências relacionadas ao processo judicial.

A partir disso, para chegar ao peso relativo das duas hipóteses para a origem das amostra questionada e de referência, é necessário realizar a aplicação da fórmula da razão de verossimilhança, a partir daqui denominada 'LR'. A fórmula de LR representada abaixo, determina a razão entre duas probabilidades, da hipótese ser de mesma origem ou da hipótese ser de origem diferente.

$$LR = \frac{p(E | H_{so})}{p(E | H_{do})} \quad (2.1)$$

Na fórmula, a primeira probabilidade calculada é  $p(E|H_{so})$ , que corresponde a probabilidade da evidência 'E' ser de mesma origem, em inglês *same origin*. A segunda probabilidade a ser calculada é  $p(E|H_{do})$ , que corresponde a probabilidade da evidência 'E' ser de origem diferente, em inglês *different origin*.

Se o valor da LR for maior que 1, pode-se dizer que prevalece a probabilidade da hipótese de mesma origem (o suspeito é a fonte da amostra questionada). Se o valor for menor do que 1, prevalece a probabilidade da hipótese de origens diferentes (o suspeito não é a fonte da amostra questionada). Valores próximos a 1 sugerem que a probabilidade de obter a evidência observada é muito semelhante considerando as duas hipóteses quanto à sua origem, de modo que o exame não ajuda a estabelecer qual das duas hipóteses têm mais peso (MORRISON, 2010).

Desta maneira, a razão de verossimilhança quantifica a força da evidência ao comparar a probabilidade de as diferenças observadas entre as amostras ocorrerem sob a hipótese de mesma origem ou de origem diferente. Podemos dizer que o numerador da LR representa a similaridade entre as amostras e seu denominador representa a tipicidade, ou seja, quão prevalente é o traço fonético observado na amostra do suspeito em relação à população relevante.

A utilização de estatísticas robustas da variabilidade populacional permite que a comparação entre amostras de origem questionada e de origem conhecida seja conduzida com maior rigor estatístico e com redução de incertezas relacionadas as probabilidades.

## 2.2 Aplicação da LR

Nesta seção, são apresentadas três simulações que ilustram o procedimento do cálculo da razão de verossimilhança. Os gráficos foram elaborados a partir dos próprios valores estatísticos obtidos neste trabalho. Essa utilização direta dos dados permite mostrar de que maneira os dados gerados por este trabalho podem contribuir em futuros exames forenses. A elaboração dos gráficos (Anexo B), foi realizada no ambiente de programação estatística R (R Core Team, 2024).

Nos três gráficos gerados, a linha vermelha representa o modelo de distribuição do parâmetro considerado na população de referência, isto é, conjunto de pessoas cujas vozes estão contempladas no acervo Projeto SP2010. A linha azul representa o valor do modelo suspeito, voz que ainda não se tem conhecimento de quem é o autor. A linha verde representa o valor obtido a partir da gravação questionada.

### CENÁRIO 1

A situação representada no cenário 1 é a da evidência favorecer origens diferentes, isto se dá pelo motivo do valor da LR encontrado ser menor do que 1 e como mostrado no gráfico, o

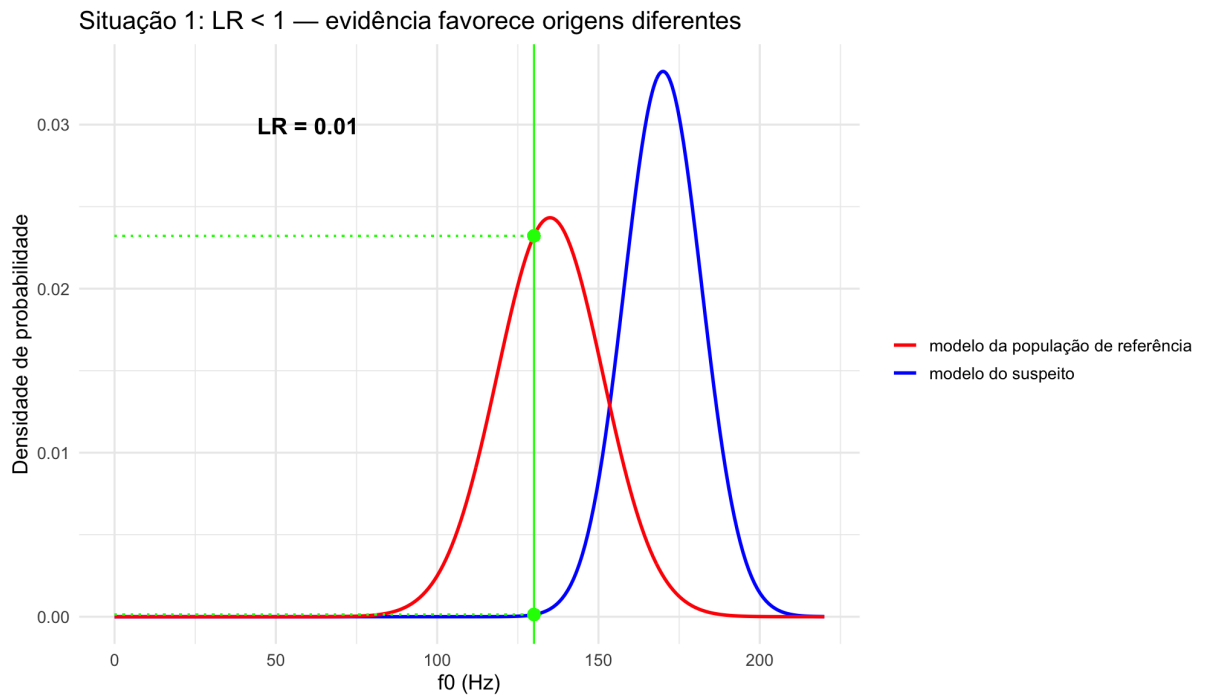


Figura 1: Gráfico de simulação do cálculo da razão de verossimilhança.

valor obtido a partir da amostra questionada abrange a curva do modelo suspeito, no entanto, a sobreposição entre as duas não é muito grande.

## CENÁRIO 2

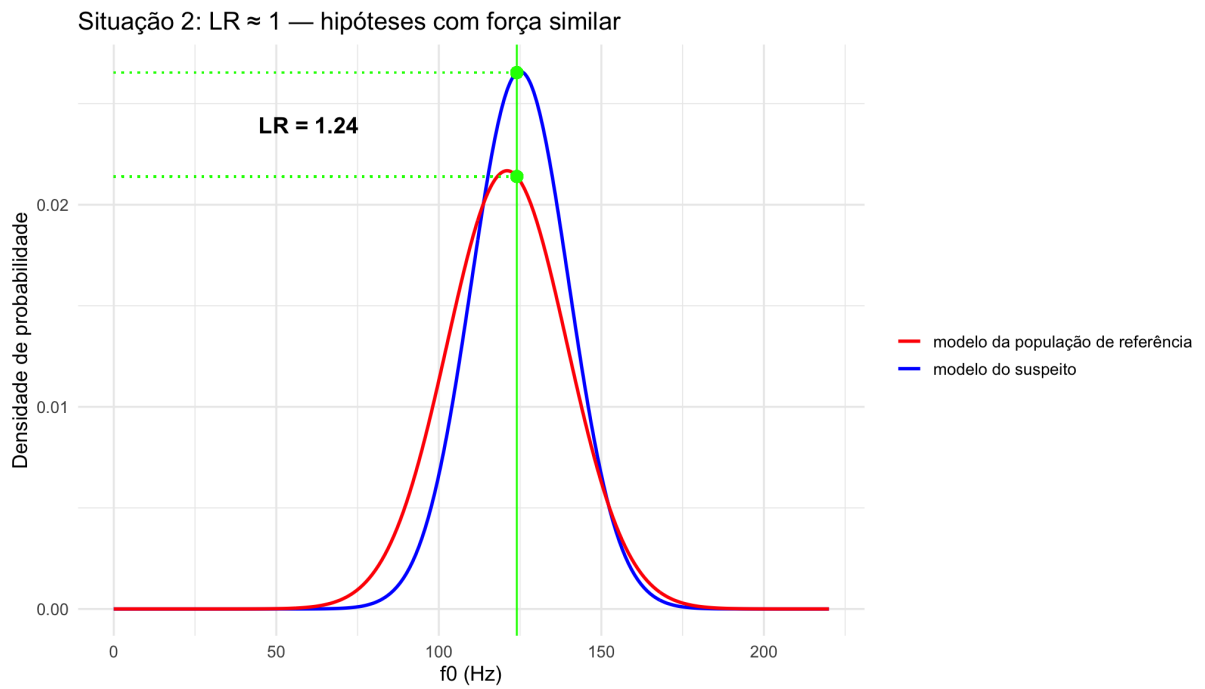


Figura 2: Gráfico de simulação do cálculo da razão de verossimilhança.

A situação representada no cenário 2 é a da hipótese de forças similares, devido ao valor de LR ser muito próximo a 1, a evidência é inconclusiva, isto é, não aponta com mais força nem para a hipótese de mesma origem nem para a de origens diferentes

### CENÁRIO 3

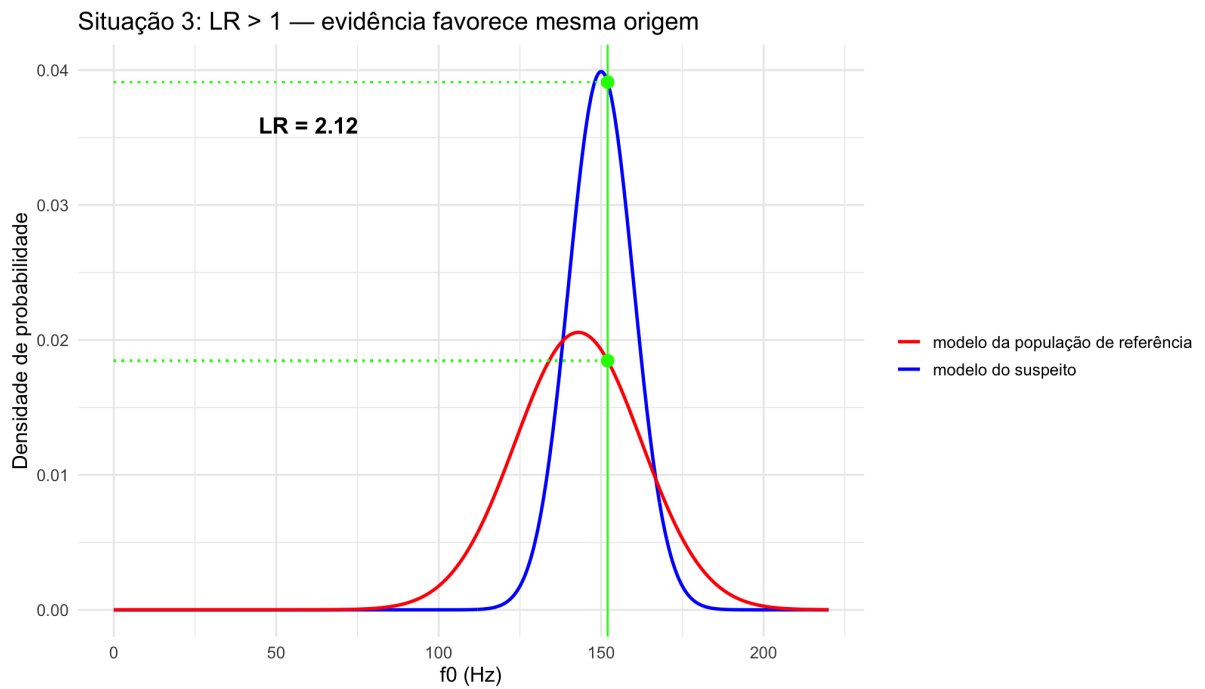


Figura 3: Gráfico de simulação do cálculo da razão de verossimilhança.

Por fim, o cenário 3 aponta para evidência favorecer a mesma origem, devido ao valor de LR ser maior do que 1, isto é, o autor da amostra questionada aponta para o mesmo indivíduo autor da amostra de referência.

## **3 METODOLOGIA**

### **3.1 Sobre o corpus utilizado**

O material de fala utilizado neste estudo provém do acervo Projeto SP2010 – “Construção de uma amostra da fala paulistana”, desenvolvido pelo Grupo de Estudos e Pesquisa em Sociolinguística da Universidade de São Paulo (GESOL-USP). A criação desse acervo representa uma contribuição significativa para os estudos sociolinguísticos brasileiros, uma vez que busca preencher a ausência de pesquisas relacionadas ao português falado na cidade de São Paulo. Diferentemente de outras regiões do país, que já dispunham de bancos de dados sociolinguísticos consolidados, a maior metrópole brasileira carecia de um acervo estratificado de sua variedade linguística. Assim, o Projeto SP2010 tem como objetivo central construir uma amostra representativa e contemporânea da fala de indivíduos nascidos e criados na cidade de São Paulo (MENDES, 2011).

Para tanto, entre os anos de 2010 e 2012, foram realizadas entrevistas com 30 falantes homens e 30 falantes mulheres. As entrevistas sociolinguísticas eram divididas em três estilos de elocução: fala espontânea, leitura de um parágrafo retirado de uma notícia relacionada à cidade de São Paulo e, por fim, leitura de palavras isoladas (MENDES, 2013).

O Projeto SP2010 estratificou os falantes por idade, escolaridade, zona geográfica, renda individual, renda familiar e geração familiar, esta última informação dizendo respeito ao histórico familiar do informante quanto à sua origem – se os pais do informante são ou não paulistanos, se os avós são ou não paulistanos (e.g.: avó se mudou para São Paulo, mãe/pai nascidos na cidade de São Paulo, falante nascido na cidade de São Paulo).

### **3.2 Seleção das amostras para análise**

Para este estudo, foram selecionadas amostras de fala de todos os falantes do sexo masculino ( $n = 30$ ), os quais eram naturais da cidade de São Paulo (SP). A escolha da análise exclusiva das vozes masculinas se deu devido à maior recorrência desse grupo em casos criminais (SILVA, 2016). Cada amostra de fala é organizada de forma a contemplar os três estilos de elocução que compõem as entrevistas do projeto SP2010: fala semiespontânea, leitura de texto e leitura de palavras isoladas. A fala espontânea, no contexto das amostras, corresponde a um diálogo entre entrevistador e entrevistado, no qual são abordados temas pessoais e cotidianos buscando

reduzir a atenção do falante à própria fala, como infância, trajetória profissional, percepção sobre o bairro de residência e sobre a cidade de São Paulo. Em seguida, o participante realiza a leitura de um parágrafo retirado de uma notícia, também relacionada à cidade de São Paulo, de modo a garantir certa uniformidade temática entre os textos lidos. Por fim, cada entrevistado lê uma lista de palavras isoladas previamente selecionadas, permitindo a análise de fala mais monitorada. As amostras possuem um tempo médio de duração de 60 minutos.

As informações sobre os falantes da amostra foram estratificadas segundo as variáveis idade, zona geográfica e escolaridade, a fim de verificar possíveis efeitos dessas variáveis sociodemográficas sobre a distribuição de  $f_0$  na população masculina paulistana. Além disso, foram geradas distribuições populacionais de diferentes estimadores estatísticos, as quais podem servir como recurso de apoio à aplicação da abordagem baseada na razão de verossimilhança em exames de CL, contribuindo para a atualização e o aperfeiçoamento da prática forense brasileira.

Para a extração da  $f_0$ , foram selecionados trechos de fala de cada estilo de elocução presente nas gravações: fala semiespontânea, leitura de frases e leitura de palavras. Essa abordagem possibilita observar a variação e distribuição dos valores de  $f_0$  em diferentes graus de monitoramento, permitindo uma análise abrangente da frequência fundamental.

### 3.3 Variáveis analisadas e métodos de extração

Os descritores estatísticos de  $f_0$  escolhidos foram: média, mediana, moda, desvio-padrão, coeficiente de variação, assimetria, curtose e valor de base. Esses parâmetros permitem uma caracterização ampla e detalhada da distribuição da  $f_0$ , contemplando medidas centrais, de dispersão e assimetria. A aplicação dessa metodologia fornece suporte técnico e científico aos peritos, estabelecendo valores de referência que podem contribuir para a precisão de investigações forenses futuras.

A extração dos dados foi realizada com o auxílio de três *scripts*: *slice merge* (Anexo A), *Better f0* (ARANTES, 2019) e *f0 outliers* (ARANTES, 2021). O *slice merge* reúne os trechos selecionados em um novo arquivo TextGrid no software Praat; o *Better f0* extrai o contorno de  $f_0$ , gerando um relatório com métricas como frequência mínima e máxima e um arquivo Pitch; já o *f0 outliers* identifica pontos que desviam significativamente do padrão de  $f_0$  do falante em curto intervalo de tempo. A inspeção dos áudios por meio desses *scripts* permitiu o descarte de alguns áudios, devido à baixa qualidade acústica, o que poderia comprometer a precisão da extração dos descritores estatísticos de  $f_0$  e levar a erros de medida. Dentre os áudios descartados, 5 correspondiam ao estilo de fala semiespontânea, 4 ao estilo de leitura de frases e 2 ao estilo de leitura de palavras.

A etapa seguinte consistiu na correção manual das curvas de  $f_0$  no Praat, a partir dos arquivos Pitch gerados pelo *Better f0*. Nesta fase é realizado um procedimento que exige investimento considerável de tempo, pois cada trecho apontado como suspeito pelo *f0 outliers* precisa ser examinado cuidadosamente. Essa verificação é fundamental para assegurar a qualidade dos dados

de maneira a evitar que as distribuições de  $f_0$  sejam distorcidas pela presença de valores espúrios decorrentes de erros de extração, ao mesmo tempo em que garante a preservação de valores atípicos legítimos, resultantes da variação natural da produção da fala. É importante destacar que algumas amostras de fala precisaram ser descartadas devido à baixa qualidade do áudio, uma vez que poderiam comprometer a precisão dos descritores estatísticos e, conseqüentemente, a confiabilidade dos resultados.

Após a correção dos trechos suspeitos apontados pelo script, o script *pitch to table* (Anexo C) foi utilizado para converter os arquivos do formato Pitch nativo do Praat em tabelas que listam os valores de  $f_0$  no formato texto puro separado por tabulação, para posterior importação no ambiente de programação estatística R (R Core Team, 2024), onde foram gerados sumários estatísticos em forma de tabelas e gráficos de diagramas de caixa para a visualização da distribuição dos diversos descritores estatísticos.

### 3.4 Análise dos dados

O trabalho iniciou-se com a seleção das amostras de áudio do acervo do projeto SP2010, seguida da segmentação conforme os estilos de elocução. Em seguida, procedeu-se à extração automática das curvas de  $f_0$  e, por fim, à correção manual desses contornos.

Os arquivos de áudio, armazenados em formato WAV, foram analisados e segmentados utilizando o software Praat (BOERSMA, 2001), ferramenta especializada em análises acústicas de fala. Através deste programa, foi possível realizar as análises acústicas detalhadas, como a extração da curva de  $f_0$  e correção de valores espúrios resultantes de erros de estimação.

Como passo inicial nas análises das amostras de falas, foram criadas 3 camadas em arquivos de anotação do tipo TextGrid que acompanham os áudios. Cada uma destas camadas foi destinada a marcações dos trechos de falas. A primeira camada, denominada “semi”, foi destinada à segmentação de trechos de fala no estilo semiespontâneo; a segunda camada, denominada “frases”, foi destinada à segmentação da leitura de trechos no estilo leitura de frases; por fim, a terceira camada, denominada “palavras”, foi destinada à segmentação da leitura de palavras isoladas.

A fala semiespontânea apresenta menor grau de monitoramento linguístico por parte do falante, o que aumenta a ocorrência de fenômenos como hesitações, disfluências e elementos paralinguísticos. Já a leitura de frases extraídas de textos jornalísticos e a leitura de palavras isoladas configuram situações mais controladas, com maior monitoramento da produção linguística. Essa distinção é fundamental para analisar padrões entoacionais e de fluência em diferentes contextos de fala, permitindo uma visão abrangente da variação dos parâmetros de  $f_0$ .

Para cada estilo de elocução, foram selecionados trechos de aproximadamente 45 segundos. Esse recorte temporal baseia-se na literatura sobre análise de  $f_0$  e outros parâmetros prosódicos, como qualidade vocal (BECK, 2007), a qual indica ser este um intervalo suficiente para capturar uma amostra estável e representativa da produção vocal, sem influência de variações

momentâneas e eventuais instabilidades prosódicas (ARANTES, 2014).

### 3.5 Análise estatística dos valores de f0

Foram gerados sumários estatísticos a partir dos contornos de f0 e aplicados testes estatísticos para entender os efeitos que as variáveis escolhidas na seção anterior têm sobre os descritores estatísticos. Além disso, foram conduzidos testes estatísticos para verificar a significância das diferenças observadas entre os grupos ou condições analisadas.

Primeiro, para que se entenda a caracterização da distribuição dos valores gerados de f0, deve-se haver uma breve explicação de como funcionam os descritores de tendência central: Moda, Mediana, Média; Os descritores de medidas de dispersão: Valor de Base, Desvio-Padrão e Coeficiente de Variação. Por último, os descritores de forma de distribuição de dados: Assimetria e Curtose. Essas medidas quando analisadas juntas, conseguem caracterizar de forma abrangente as distribuições de f0.

A Média é a média aritmética da amostra, calculada pela soma de todas as ocorrências dividida pelo número total de amostras.

A Mediana corresponde ao valor central da distribuição de f0, este estimador é definido através de dois passos, primeiro a amostra é ordenada do menor para o maior valor. Depois o conjunto de dados é segmentado em duas porções de tamanhos iguais. Quando o tamanho da amostra resulta em dois termos centrais, a mediana é determinada pelo cálculo da média aritmética entre esses dois valores. Ela é considerada mais robusta que a média na presença de valores *outliers*.

A Moda indica o valor de f0 que ocorre com maior frequência no conjunto de dados e é calculada a partir da aplicação da técnica *Kernel Density Estimation*. Ao realizar esse cálculo, estima-se a densidade de probabilidade da amostra dos dados totais de f0 através de uma versão suavizada de um histograma, que apresenta uma curva contínua. Além disso, o valor presente no ponto máximo global dessa curva pode ser considerado o valor mais comum na amostra (CUNHA, 2023).

O Valor de base é considerado um estimador estatístico do valor de f0, baseado nos valores da média e desvio padrão. O cálculo que é feito especificamente para encontrar o valor de f0 é feito com base na diferença entre a média aritmética e a multiplicação da constante  $k$ , que é representada pela fórmula:

$$F_b = F_{\text{média}} - k\sigma$$

O Desvio-Padrão quantifica o grau de dispersão dos valores de f0 em torno da média; o Coeficiente de Variação é uma medida de dispersão relativa, que é expressa em porcentagem e indica o tamanho do desvio-padrão em relação à média, permitindo comparar a variabilidade de conjuntos de dados com médias diferentes.

A Assimetria descreve a simetria da distribuição dos valores de  $f_0$ , um valor próximo de zero sugere uma distribuição simétrica, enquanto, uma assimetria à direita é considerada positiva, que significa uma maior concentração de valores baixos, enquanto uma assimetria à esquerda é considerada negativa apontando para maior concentração de valores altos; a Curtose é usada para complementar outras medidas estatísticas, como a média e o desvio-padrão, para obter uma compreensão mais completa da distribuição dos dados, ela mede o grau de achatamento da curva de distribuição quando comparado com a distribuição normal.

Os descritores estatísticos foram calculadas para os valores de frequência fundamental ( $f_0$ ) considerando duas escalas distintas: a escala física, expressa em Hertz (Hz), e a escala relativa, expressa em semitons (st). A utilização dessas duas escalas permite observar a variação da  $f_0$  tanto em termos absolutos, relacionados à medida física da frequência, quanto em termos perceptivos, de forma a possibilitar uma análise mais abrangente dos padrões vocais entre as amostras dos falantes.

Posteriormente, para a análise visual da distribuição dos dados, foram gerados gráficos de diagramas de caixa. Este tipo de gráfico é eficaz para representar a distribuição, exibindo a mediana, consistindo de uma linha central, os quartis, que são as "bordas" da caixa e, a presença de possíveis *outliers*, que são representados por pontos individuais além das caixas.

### 3.6 Testes estatísticos

Após a análise dos valores encontrados, foram realizados testes estatísticos para verificar a presença de significância nas diferenças observadas entre os valores gerados por cada grupo. O teste de homogeneidade de variância, chamado de *Fligner-Killeen*, foi aplicado em primeiro lugar para determinar se as variâncias dos grupos eram homogêneas ou não. Para que se determine essa variância, há duas possibilidades de resultados: se o valor de  $p$  for maior que 0,05 ( $p > 0,05$ ), a hipótese de homogeneidade das variâncias é aceita, indicando que os grupos possuem dispersões similares. A outra possibilidade é, caso o valor de  $p$  for menor ou igual a 0,05 ( $p < 0,05$ ), a hipótese de homogeneidade é rejeitada, indicando variâncias heterogêneas.

A partir dos valores de  $p$ , têm-se duas possibilidades de testes, o teste não-paramétrico *Kruskal-Wallis* e o teste paramétrico Análise de Variância, ou *ANOVA*. O primeiro, chamado de *Kruskal-Wallis*, é realizado caso o valor de  $p$  seja menor que 0,05, este teste compara as medianas dos grupos, sendo mais robusto para dados que não atendem aos pressupostos de normalidade ou homocedasticidade. O segundo teste, chamado de *ANOVA*, é feito caso o valor de  $p$  seja maior que 0,05, e ele é usado para testar as diferenças entre duas ou mais médias. Sua função é mostrar se existe diferença significativa entre os valores dos grupos que estão sendo comparados.

Caso aconteça de estes dois testes indicarem uma diferença significativa geral entre os grupos, tornava-se necessário identificar quais pares de grupos específicos diferem entre si. Para isso, aplica-se um teste *post-hoc*, neste trabalho, foram utilizados os testes  $t$ , para quando a *ANOVA* foi realizada anteriormente e o teste de *Wilcoxon*, que é um teste não-paramétrico utilizado para com-

parar dois grupos pareados ou relacionados, permitindo encontrar as diferenças estatisticamente significantes que foram apontadas pelo teste não paramétrico realizado anteriormente.

## 4 RESULTADOS E DISCUSSÃO

A partir dos dados estruturados com as tabelas, foi possível chegar ao objetivo final deste trabalho, apresentando os valores de f0 nos diferentes descritores estatísticos. Para além da representação dos resultados em forma de tabelas, optou-se por uma representação visual mais clara e comparativa das distribuições de f0 através da utilização dos diagramas de caixa, gerados no software R.

As Tabelas detalham os valores da média e do desvio-padrão (DP) para um conjunto de descritores estatísticos. Elas estão organizadas em três categorias principais: medidas de tendência central (moda, mediana, média e valor de base), medidas de dispersão (desvio-padrão e coeficiente de variação) e medidas de forma de distribuição (assimetria e curtose).

### 4.1 Efeito da escolaridade sobre os descritores estatísticos

Os valores na tabela 1 são em função da variável escolaridade, que compreende dois níveis: *até médio*, para falantes com escolaridade até o ensino médio completo, e *superior*, para informantes com ensino superior completo. Os valores são apresentados nas unidades de medida: Hertz (Hz) e Semitons (ST). A coluna "N" corresponde ao número de amostras utilizadas no cálculo de cada métrica, sendo 36 para o grupo "até médio" e 37 para o grupo "superior".

Tabela 1: Medidas de f0 em Hertz (Hz) e Semitons (ST), por nível de Escolaridade.

| Métrica | Grupo     | Valores em Semitons |      | Valores em Hertz |      | N  |
|---------|-----------|---------------------|------|------------------|------|----|
|         |           | Média               | DP   | Média            | DP   |    |
| Moda    | até médio | 83,7                | 2,41 | 125              | 18,4 | 36 |
|         | superior  | 83,9                | 3,79 | 130              | 30,5 | 37 |
| Média   | até médio | 84,0                | 2,28 | 131              | 17,2 | 36 |
|         | superior  | 84,3                | 3,35 | 134              | 27,5 | 37 |
| Mediana | até médio | 83,9                | 2,31 | 128              | 17,2 | 36 |
|         | superior  | 84,2                | 3,43 | 132              | 27,8 | 37 |

Medidas de f0 em Hertz (Hz) e Semitons (ST), por Nível de Escolaridade.

| Métrica                 | Grupo     | Valores em Semitons |      | Valores em Hertz |      | N  |
|-------------------------|-----------|---------------------|------|------------------|------|----|
|                         |           | Média               | DP   | Média            | DP   |    |
| Valor de base           | até médio | 80,8                | 2,98 | 108              | 17,7 | 36 |
|                         | superior  | 81,2                | 3,23 | 111              | 21,8 | 37 |
| Desvio Padrão           | até médio | 2,39                | 0,96 | 18,4             | 6,81 | 36 |
|                         | superior  | 2,35                | 0,56 | 18,7             | 5,85 | 37 |
| Coeficiente de variação | até médio | 2,86                | 1,17 | 14,2             | 5,69 | 36 |
|                         | superior  | 2,80                | 0,68 | 13,9             | 3,35 | 37 |
| Assimetria              | até médio | 0,29                | 0,50 | 0,72             | 0,53 | 36 |
|                         | superior  | 0,25                | 0,41 | 0,64             | 0,45 | 37 |
| Curtose                 | até médio | 3,50                | 0,91 | 4,11             | 1,50 | 36 |
|                         | superior  | 3,18                | 0,66 | 3,61             | 1,19 | 37 |

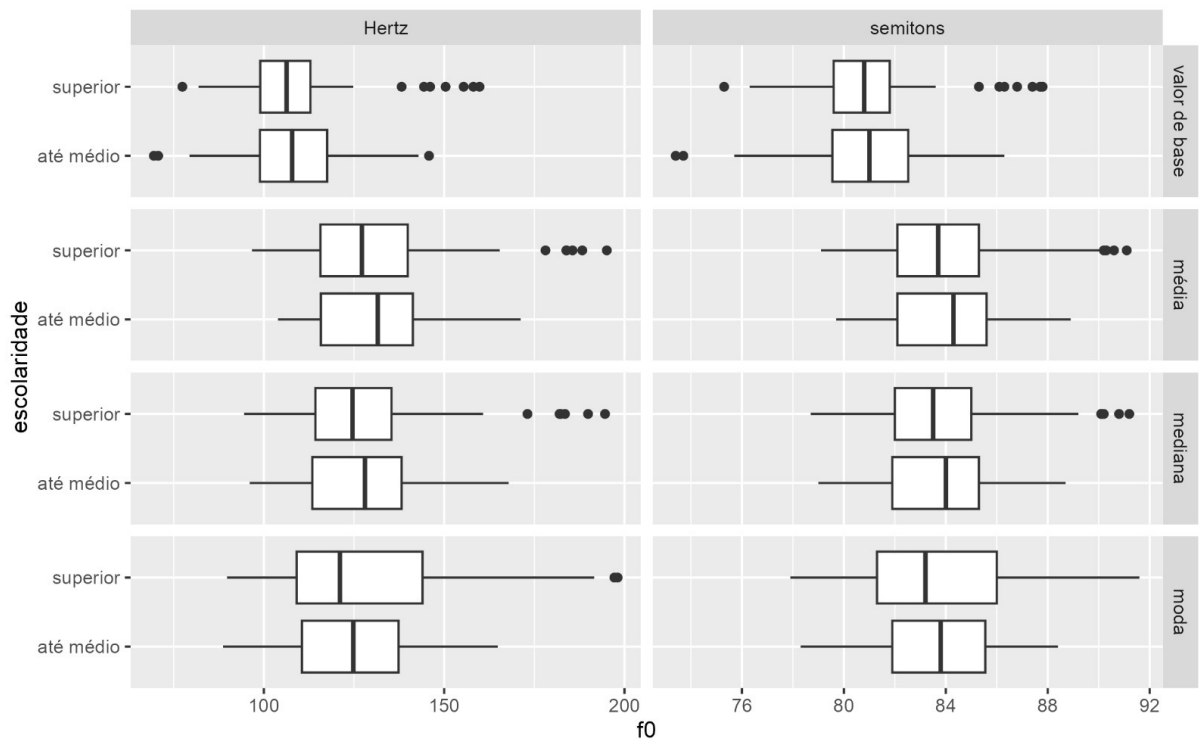


Figura 4: Diagramas de caixa dos descritores de tendência central f0 em função da escala de mensuração e do nível de escolaridade dos participantes.

Ao analisar as medidas de tendência central, observa-se um padrão consistente: o grupo com escolaridade "superior" apresenta valores médios de  $f_0$  ligeiramente elevados em todas as métricas quando comparado ao grupo "até médio". A média da  $f_0$  para o grupo "superior" foi de 134 Hz, enquanto para o grupo "até médio" foi de 131 Hz. O mesmo padrão é verificado para a moda, 130 Hz (superior) e 125 Hz (até médio); mediana: 132 Hz (superior) e 128 Hz (até médio); valor de base: 111 Hz (superior) e 108 Hz (até médio). Este resultado indica que os falantes com maior nível de escolaridade tendem a apresentar uma  $f_0$  ligeiramente mais alta. Além disso, o desvio-padrão associado a essas medidas foi maior para o grupo "superior", com a média sendo 27,5 Hz (superior) e 17,2 Hz (até médio), o que indica uma variabilidade maior entre os indivíduos com o ensino superior completo.

No que tange às medidas de dispersão, que refletem a variabilidade da  $f_0$ , os resultados mostram uma notável semelhança entre os grupos. O desvio-padrão médio da  $f_0$  foi semelhante, com 18,7 Hz para o grupo "superior" e 18,4 Hz para o grupo "até médio". De maneira semelhante, o coeficiente de variação, que mede a dispersão relativa, também se mostrou muito próximo entre os dois grupos, com valores de 13,9 Hz para "superior" e 14,2 Hz para "até médio".

Por fim, ao analisar as medidas de forma da distribuição, observou-se que em ambos os grupos apresentaram uma leve inclinação para valores mais baixos de frequência: 0,64 Hz (superior) e 0,72 Hz (até médio). Isso quer dizer que a maioria dos valores de  $f_0$  se concentram nas frequências mais baixas, com algumas ocorrências chegando a frequências mais altas. Quanto à curtose, ambos os grupos mostraram valores acima de 3 — 3,61 Hz (superior) e 4,11 Hz (até médio), isso indica que os resultados estão mais concentrados em torno da média, mas ainda assim aparecem alguns valores *outliers*.

O teste de *Fligner-Killeen* foi conduzido para verificar a homogeneidade de variância entre esses dois grupos. Os valores encontrados foram: Moda (ST):  $p = 0,07318$ ; Média (ST):  $p = 0,3112$ ; Mediana (ST):  $p = 0,2103$ ; Valor de base (ST):  $p = 0,7587$ .

Apesar de falantes com formação superior completa apresentarem uma  $f_0$  ligeiramente mais alta do que falantes com ensino médio completo, a partir desses resultados, conclui-se que os dois níveis de escolaridade não exercem nenhum efeito sobre a média dos estimadores de  $f_0$ , não havendo necessidade de aplicação de testes *post-hoc*. No entanto, a presença de *outliers* indica que existem alguns indivíduos em ambos os grupos com frequências consideravelmente mais altas do que o padrão apresentado pela maioria dos falantes.

## 4.2 Efeito dos estilos de elocução sobre os descritores estatísticos

Os valores apresentados na tabela 2 estão em função da variável estilos de elocução, que consistem: fala semiespontânea, leitura de frases e leitura de palavras. A coluna "N" informa o número de amostras utilizadas para o cálculo em cada condição, sendo 25 para a fala semiespontânea, 22 para a leitura de frases e 26 para a leitura de palavras.

Tabela 2: Medidas de f0 em Hertz (Hz) e Semitons (ST) por Estilo de Elocução.

| Métrica                 | Estilo   | Valores em Semitons |      | Valores em Hertz |      | N  |
|-------------------------|----------|---------------------|------|------------------|------|----|
|                         |          | Média               | DP   | Média            | DP   |    |
| Moda                    | semi     | 83,1                | 3,32 | 123              | 24,8 | 25 |
|                         | frases   | 84,1                | 2,98 | 129              | 24,4 | 22 |
|                         | palavras | 84,3                | 3,16 | 131              | 26,7 | 26 |
| Média                   | semi     | 84,0                | 2,90 | 132              | 22,7 | 25 |
|                         | frases   | 84,4                | 2,86 | 134              | 23,0 | 22 |
|                         | palavras | 84,2                | 2,90 | 132              | 23,8 | 26 |
| Mediana                 | semi     | 83,7                | 3,06 | 128              | 23,6 | 25 |
|                         | frases   | 84,2                | 2,88 | 131              | 22,9 | 22 |
|                         | palavras | 84,1                | 2,91 | 131              | 23,6 | 26 |
| Valor de base           | semi     | 80,3                | 3,42 | 105              | 20,4 | 25 |
|                         | frases   | 81,2                | 2,98 | 110              | 19,6 | 22 |
|                         | palavras | 81,6                | 2,82 | 113              | 19,5 | 26 |
| Desvio Padrão           | semi     | 2,76                | 0,76 | 21,1             | 5,20 | 25 |
|                         | frases   | 2,47                | 0,71 | 19,8             | 6,13 | 22 |
|                         | palavras | 1,91                | 0,61 | 15,0             | 5,98 | 26 |
| Coeficiente de variação | semi     | 3,30                | 0,96 | 16,3             | 4,38 | 25 |
|                         | frases   | 2,95                | 0,86 | 14,8             | 4,35 | 22 |
|                         | palavras | 2,28                | 0,72 | 11,2             | 3,69 | 26 |
| Assimetria              | semi     | 0,24                | 0,45 | 0,66             | 0,45 | 25 |
|                         | frases   | 0,34                | 0,41 | 0,84             | 0,42 | 22 |
|                         | palavras | 0,23                | 0,51 | 0,57             | 0,55 | 26 |

Medidas de f0 em Hertz (Hz) e Semitons (ST) por Estilo de Elocução.

| Métrica | Estilo   | Valores em Semitons |      | Valores em Hertz |      | N  |
|---------|----------|---------------------|------|------------------|------|----|
|         |          | Média               | DP   | Média            | DP   |    |
| Curtose | semi     | 3,07                | 0,60 | 3,43             | 1,06 | 25 |
|         | frases   | 3,58                | 0,71 | 4,42             | 1,42 | 22 |
|         | palavras | 3,38                | 0,98 | 3,79             | 1,45 | 26 |

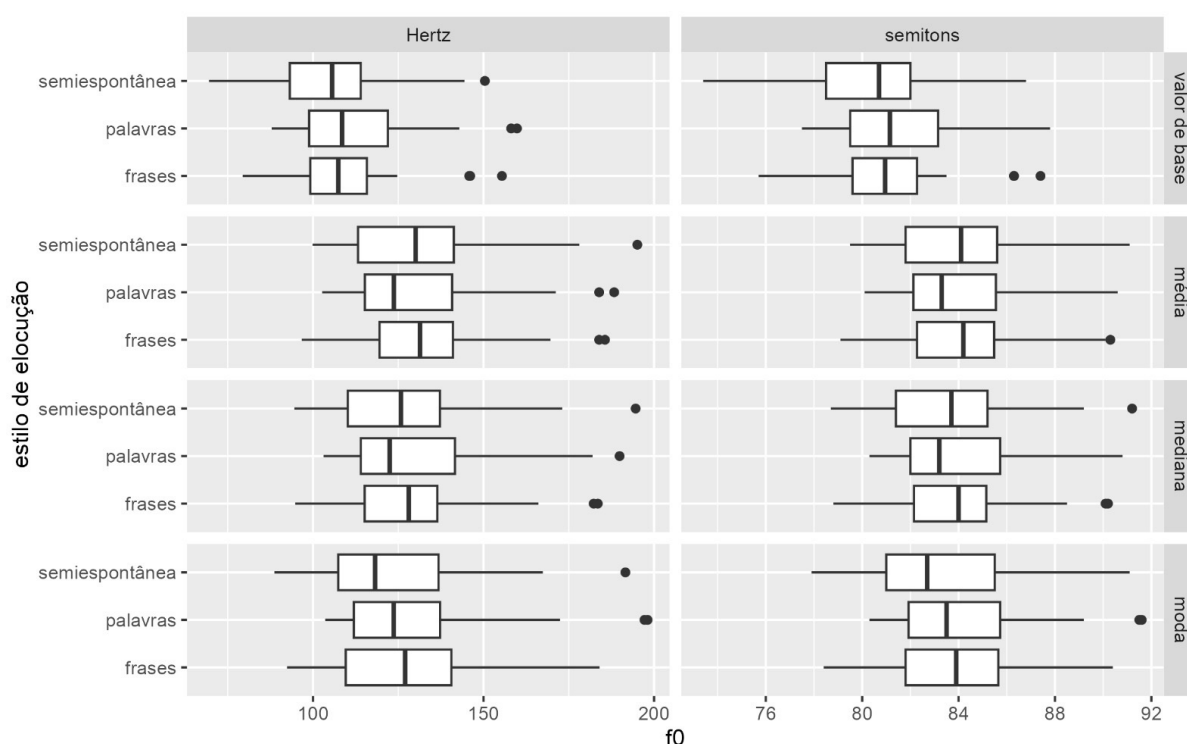


Figura 5: Diagramas de caixa dos descritores de tendência central f0 em função da escala de mensuração e dos estilos de elocução dos participantes.

Ao analisar as medidas de tendência central, é possível observar uma certa similaridade nos valores de f0 entre os diferentes estilos. A média, permanece bastante estável, com valores de 132 Hz para os estilos de fala semiespontânea e leitura de palavras, e 134 Hz para a leitura de frases.

Um padrão similar é observado na mediana e na moda, sugerindo que o tom médio dos falantes não sofreu alterações em função do estilo de elocução. O valor de base apresentou um pequeno aumento, partindo de 105 Hz no estilo de fala semiespontânea, passando para 110 Hz na leitura de frases e por fim, 113 Hz na leitura de palavras. Esses valores mostram um certo aumento de f0 em situações de leitura controlada.

Ao contrário dos descritores de tendência central, as medidas de dispersão apontam diferenças entre os estilos de elocução. O desvio-padrão e o coeficiente de variação, mostram uma tendência decrescente respectivamente aos estilos de elocução. A fala semiespontânea apresenta os

seguintes valores: Desvio-padrão = 21,1 Hz e Coeficiente de variação = 16,3 Hz; o estilo leitura de frases apresenta valores: Desvio-padrão = 19,8 Hz e Coeficiente de variação = 14,8 Hz e, por fim, o estilo leitura de palavras apresenta valores: Desvio-padrão = 15,0 Hz e Coeficiente de variação = 11,2Hz. Este padrão sugere que a fala espontânea apresenta uma variabilidade maior, quando comparado com estilos de fala controlada, como os estilos de leitura de frases e palavras.

Ao observar os valores de assimetria, os três estilos mostraram valores de assimetria positivos. Para os valores da curtose, todos os estilos de elocução apresentaram valores maiores que 3, indicando resultados mais concentrados em torno da média, algo a ser destacado é o valor apresentado pelo estilo de leitura de frases, com o maior valor da curtose, 4,42 Hz.

Além disso, ao analisar a dimensão da caixa de  $f_0$  que representa o estilo de fala semiespontânea, mostra que há presença de variações na fala, devido esse registro permitir um menor grau de monitoramento pelo falante, aumentando a chance de surgir fenômenos como hesitações, disfluências e manifestação de fatores paralinguísticos. Essas características são menos frequentes em situações de fala mais controlada, como o estilo de leitura de palavras, cujas amostras são produzidas com maior grau de monitoramento linguístico por parte dos falantes.

Através dos testes de *Fligner-Killeen*, os valores encontrados para os descritores foram: Moda (ST):  $p = 0,6457$ ; Média (ST):  $p = 0,9535$ ; Mediana (ST):  $p = 0,9806$ ; Valor de base (ST):  $p = 0,8174$ . Desta maneira, após observar os valores gerados, conclui-se que os estilos de elocução não exercem nenhum efeito sobre a média dos estimadores de tendência central.

Os valores encontrados nessa variável se assemelham aos valores apresentados no estudo realizado anteriormente por Suardi (2023), com a moda da fala espontânea e leitura sendo respectivamente 122,1 Hz e 131,1 Hz; além disso a mediana também apresentou valores parecidos, com 129,9 Hz (espontânea) e 131,8 Hz (leitura).

### **4.3 Efeito das zonas geográficas da cidade de São Paulo sobre os descritores estatísticos**

Os valores apresentados na tabela 3 estão em função da variável zonas geográficas da cidade de São Paulo, que consistem das zonas: Centro, Norte, Leste, Oeste e Sul. A coluna "N" informa o número de amostras utilizadas para o cálculo, consistindo de 15 (Centro), 18 (Norte), 14 (Leste), 16 (Oeste) e 10 (Sul).

Tabela 3: Medidas de f0 em Hertz (Hz) e Semitons (ST) por Zona Geográfica.

| Métrica                 | Zona   | Valores em Semitons |      | Valores em Hertz |      | N  |
|-------------------------|--------|---------------------|------|------------------|------|----|
|                         |        | Média               | DP   | Média            | DP   |    |
| Moda                    | Centro | 82,0                | 1,71 | 111              | 11,5 | 15 |
|                         | Norte  | 84,4                | 4,11 | 134              | 33,4 | 18 |
|                         | Leste  | 84,5                | 3,53 | 134              | 26,9 | 14 |
|                         | Oeste  | 85,0                | 2,60 | 136              | 21,3 | 16 |
|                         | Sul    | 82,5                | 1,74 | 117              | 11,5 | 10 |
| Média                   | Centro | 82,7                | 1,51 | 121              | 9,89 | 15 |
|                         | Norte  | 84,9                | 3,80 | 139              | 31,2 | 18 |
|                         | Leste  | 84,9                | 3,42 | 139              | 27,7 | 14 |
|                         | Oeste  | 84,9                | 2,02 | 137              | 15,9 | 16 |
|                         | Sul    | 82,9                | 1,65 | 122              | 11,6 | 10 |
| Mediana                 | Centro | 82,5                | 1,56 | 118              | 10,3 | 15 |
|                         | Norte  | 84,7                | 3,89 | 137              | 31,8 | 18 |
|                         | Leste  | 84,8                | 3,50 | 136              | 27,7 | 14 |
|                         | Oeste  | 84,8                | 1,91 | 135              | 14,8 | 16 |
|                         | Sul    | 82,7                | 1,73 | 119              | 11,7 | 10 |
| Valor de base           | Centro | 79,4                | 2,45 | 99,1             | 13,1 | 15 |
|                         | Norte  | 82,0                | 3,53 | 117              | 24,0 | 18 |
|                         | Leste  | 81,8                | 3,62 | 115              | 24,3 | 14 |
|                         | Oeste  | 81,3                | 2,78 | 111              | 16,4 | 16 |
|                         | Sul    | 79,9                | 1,73 | 102              | 9,70 | 10 |
| Desvio Padrão           | Centro | 2,56                | 1,09 | 18,4             | 7,63 | 15 |
|                         | Norte  | 2,14                | 0,68 | 17,8             | 7,08 | 18 |
|                         | Leste  | 2,29                | 0,58 | 18,6             | 5,32 | 14 |
|                         | Oeste  | 2,56                | 0,76 | 20,5             | 5,76 | 16 |
|                         | Sul    | 2,31                | 0,65 | 17,0             | 5,06 | 10 |
| Coeficiente de variação | Centro | 3,09                | 1,34 | 15,3             | 6,76 | 15 |
|                         | Norte  | 2,54                | 0,82 | 12,7             | 4,07 | 18 |
|                         | Leste  | 2,71                | 0,73 | 13,6             | 3,67 | 14 |
|                         | Oeste  | 3,03                | 0,89 | 14,9             | 3,84 | 16 |
|                         | Sul    | 2,80                | 0,78 | 13,9             | 4,09 | 10 |

Medidas de f0 em Hertz (Hz) e Semitons (ST) por Zona Geográfica.

| Métrica    | Zona   | Valores em Semitons |      | Valores em Hertz |      | N  |
|------------|--------|---------------------|------|------------------|------|----|
|            |        | Média               | DP   | Média            | DP   |    |
| Assimetria | Centro | 0,36                | 0,34 | 0,76             | 0,44 | 15 |
|            | Norte  | 0,35                | 0,49 | 0,73             | 0,48 | 18 |
|            | Leste  | 0,17                | 0,33 | 0,56             | 0,41 | 14 |
|            | Oeste  | 0,087               | 0,58 | 0,57             | 0,60 | 16 |
|            | Sul    | 0,40                | 0,42 | 0,82             | 0,51 | 10 |
| Curtose    | Centro | 3,13                | 0,86 | 3,80             | 1,39 | 15 |
|            | Norte  | 3,50                | 0,95 | 3,96             | 1,42 | 18 |
|            | Leste  | 3,05                | 0,59 | 3,53             | 1,37 | 14 |
|            | Oeste  | 3,62                | 0,74 | 4,00             | 1,39 | 16 |
|            | Sul    | 3,30                | 0,70 | 3,98             | 1,35 | 10 |

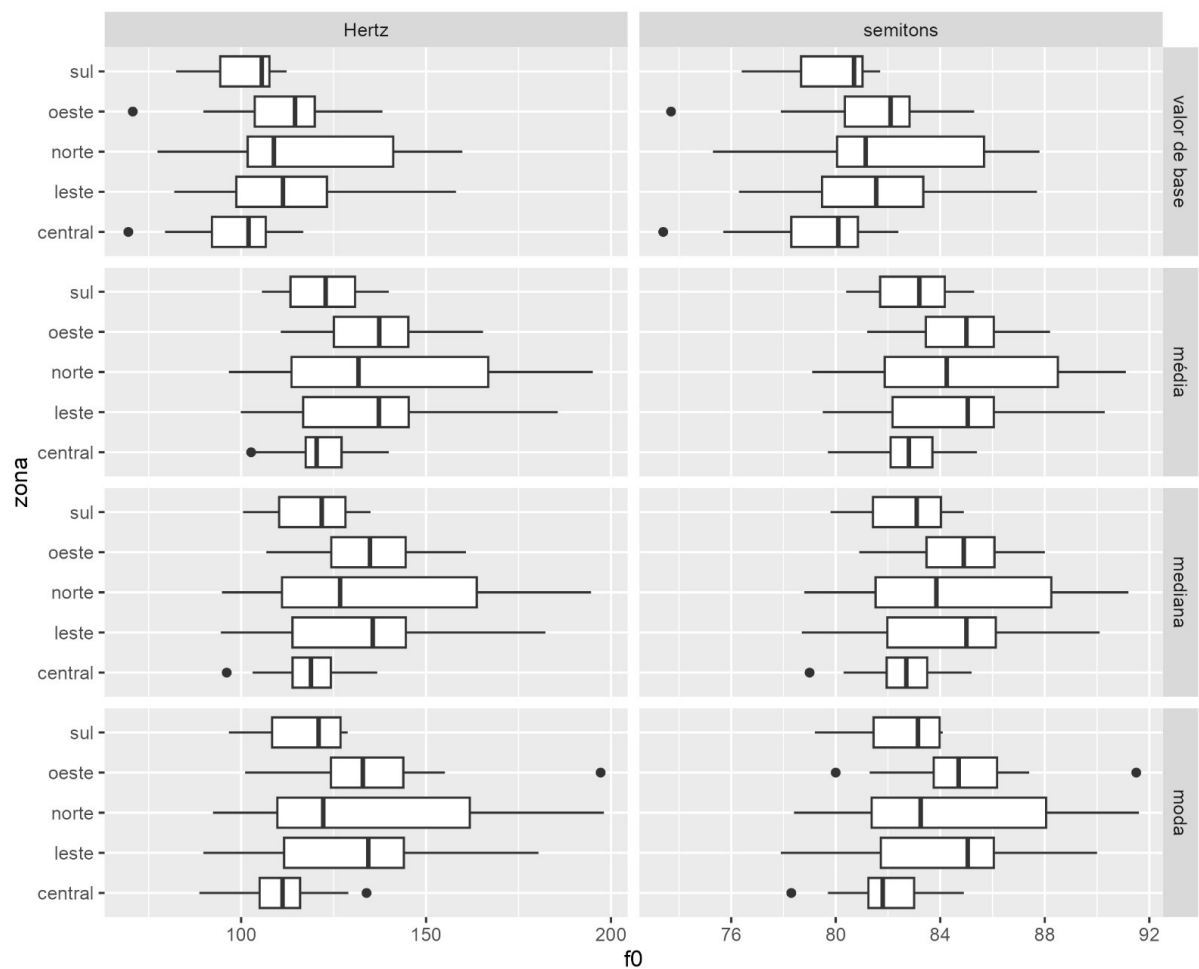


Figura 6: Diagramas de caixa dos descritores de tendência central f0 em função da escala de mensuração e das zonas geográficas da cidade de São Paulo.

Ao analisar as medidas de tendência central, é possível observar um padrão de agrupamento geográfico, com a formação de dois grupos: Zona Central e Zona Sul apresentam valores da média de  $f_0$  consideravelmente mais baixos, 121 Hz e 122 Hz, respectivamente. Enquanto isso, as zonas Norte, Leste e Oeste formam um segundo grupo com valores da média de  $f_0$  mais elevados e muito próximos entre si, sendo 139 Hz, 139 Hz e 137 Hz, respectivamente. Esse mesmo padrão se repete de forma consistente nas outras medidas de tendência central, com as zonas Central e Sul sempre apresentando valores bem abaixo do que as outras três regiões. Esses resultados sugerem uma possível diferença dialetológica entre as zonas Central e Norte.

Através do diagrama de caixa, é possível observar essa variação apresentada pelos valores da tabela, com as caixas das zonas Sul e Central tendo tamanhos parecidos, com baixa variabilidade nos descritores. Por outro lado a caixa que representa os valores dos falantes da Zona Norte apresentou a maior variabilidade de  $f_0$  entre as cinco zonas.

Em relação às medidas de dispersão, todas as regiões apresentam valores parecidos, ao contrário da clara separação observada nos descritores de tendência central. Os valores de desvio-padrão e do coeficiente de variação são semelhantes entre as cinco zonas geográficas. A média do desvio-padrão de  $f_0$  varia pouco, com valores: 18,4 Hz (Zona Centro); 17,8 Hz (Zona Norte); 18,6 Hz (Zona Leste); 17,0 Hz (Zona Sul) e 20,5 Hz (Zona Oeste). O coeficiente de variação também mostra essa semelhança, variando de 12,7 Hz na Zona Norte a 15,3 na zona Centro).

Quanto às medidas de forma de distribuição, a assimetria foi positiva em todas as zonas, os valores da curtose, todas as cinco zonas apresentaram valores acima de 3, o que aponta para distribuições mais concentradas ao redor da média.

As diferenças comentadas anteriormente são confirmadas pela análise estatística, com os testes de *Fligner-Killeen* e o teste não-paramétrico *Kruskal-Wallis* sendo aplicados. O teste de *Fligner-Killeen* apontou  $p = 0,01294$ , apresentando diferença estatisticamente significativa na dispersão da Moda (Hz), na aplicação do teste de *Kruskal-Wallis*,  $p = 0,006566$ . Como aplicação de teste *post-hoc*, o Teste de *Wilcoxon* apontou que as zonas Sul e Central têm valores estatisticamente mais baixos do que as demais zonas, com os valores sendo: Zona Sul:  $p = 0,0303$  e Zona Central:  $p = 0,0021$ .

Para a métrica da Mediana (Hz), o teste de *Fligner-Killeen* apontou  $p = 0,005478$ ; a aplicação do *Kruskal-Wallis* resultou em  $p = 0,03073$ , mostrando uma diferença estatisticamente significativa na Mediana (Hz) entre, pelo menos, duas zonas. Por fim, a aplicação de teste *post-hoc Wilcoxon* apontou diferença sistemática entre as zonas Central e Oeste:  $p = 0,012$ , e, a diferença entre as Zonas Sul e Oeste é marginalmente significativa:  $p = 0,061$ .

Por fim, a análise da distribuição etária entre as diferentes zonas geográficas revela um cenário relativamente equilibrado. As médias de idade observadas foram de 52 anos na Zona Central, 44 na Zona Sul, 46 na Zona Norte, 49 na Zona Oeste e 46 na Zona Leste. Considerando esses valores, não é evidente que o efeito estatístico associado à variável “zona geográfica” decorra de uma distribuição irregular das idades entre essas zonas.

Uma explicação alternativa é que esses valores podem estar relacionados à heterogeneidade das variâncias entre as zonas, evidenciada pelos testes de variância. Essa heterogeneidade, por sua vez, pode decorrer do número relativamente reduzido de amostras disponíveis em cada zona.

Até o momento, não há estudos em dialetologia ou sociofonética voltados especificamente para a cidade de São Paulo que permitam explicar, de forma conclusiva, as diferenças observadas neste trabalho.

#### 4.4 Efeito de diferentes faixas etárias sobre os descritores estatísticos

Os valores apresentados na tabela 4 estão em função da variável de faixa etária, e foram estratificadas pelo Projeto SP2010 em três grupos: 19-34 anos, 35-60 anos e 60+ anos. A coluna "N" informa o número de amostras utilizadas para o cálculo, 24 (19-34); 22 (35-60); 27 (60+).

Tabela 4: Medidas de f0 em Hertz (Hz) e Semitons (ST) por Faixa Etária.

| Métrica       | Faixa Etária | Valores em Semitons |      | Valores em Hz |      | N  |
|---------------|--------------|---------------------|------|---------------|------|----|
|               |              | Média               | DP   | Média         | DP   |    |
| Moda          | 19-34        | 84,2                | 2,57 | 130           | 20,3 | 24 |
|               | 35-60        | 84,9                | 3,70 | 137           | 29,7 | 22 |
|               | 60+          | 82,6                | 2,87 | 117           | 22,4 | 27 |
| Média         | 19-34        | 84,6                | 2,47 | 135           | 20,8 | 24 |
|               | 35-60        | 85,4                | 3,35 | 143           | 27,6 | 22 |
|               | 60+          | 82,8                | 2,18 | 122           | 15,8 | 27 |
| Mediana       | 19-34        | 84,4                | 2,50 | 133           | 20,6 | 24 |
|               | 35-60        | 85,2                | 3,46 | 140           | 28,2 | 22 |
|               | 60+          | 82,7                | 2,28 | 120           | 16,1 | 27 |
| Valor de base | 19-34        | 82,0                | 2,36 | 115           | 16,8 | 24 |
|               | 35-60        | 82,4                | 3,31 | 119           | 22,4 | 22 |
|               | 60+          | 79,1                | 2,51 | 97,1          | 13,5 | 27 |
| Desvio Padrão | 19-34        | 2,04                | 0,48 | 16,4          | 4,81 | 24 |
|               | 35-60        | 2,29                | 0,52 | 19,4          | 5,46 | 22 |
|               | 60+          | 2,73                | 1,00 | 19,8          | 7,66 | 27 |

Medidas de f0 em Semitons (ST) e Hertz (Hz) por Faixa Etária.

| Métrica                 | Faixa Etária | Valores em Semitons |      | Valores em Hz |      | N  |
|-------------------------|--------------|---------------------|------|---------------|------|----|
|                         |              | Média               | DP   | Média         | DP   |    |
| Coeficiente de variação | 19-34        | 2,41                | 0,57 | 12,1          | 2,79 | 24 |
|                         | 35-60        | 2,70                | 0,64 | 13,7          | 3,29 | 22 |
|                         | 60+          | 3,31                | 1,21 | 16,1          | 5,96 | 27 |
| Assimetria              | 19-34        | 0,38                | 0,43 | 0,77          | 0,43 | 24 |
|                         | 35-60        | 0,29                | 0,42 | 0,69          | 0,49 | 22 |
|                         | 60+          | 0,16                | 0,49 | 0,59          | 0,54 | 27 |
| Curtose                 | 19-34        | 3,57                | 0,96 | 4,15          | 1,49 | 24 |
|                         | 35-60        | 3,35                | 0,68 | 3,85          | 1,23 | 22 |
|                         | 60+          | 3,11                | 0,70 | 3,60          | 1,34 | 27 |

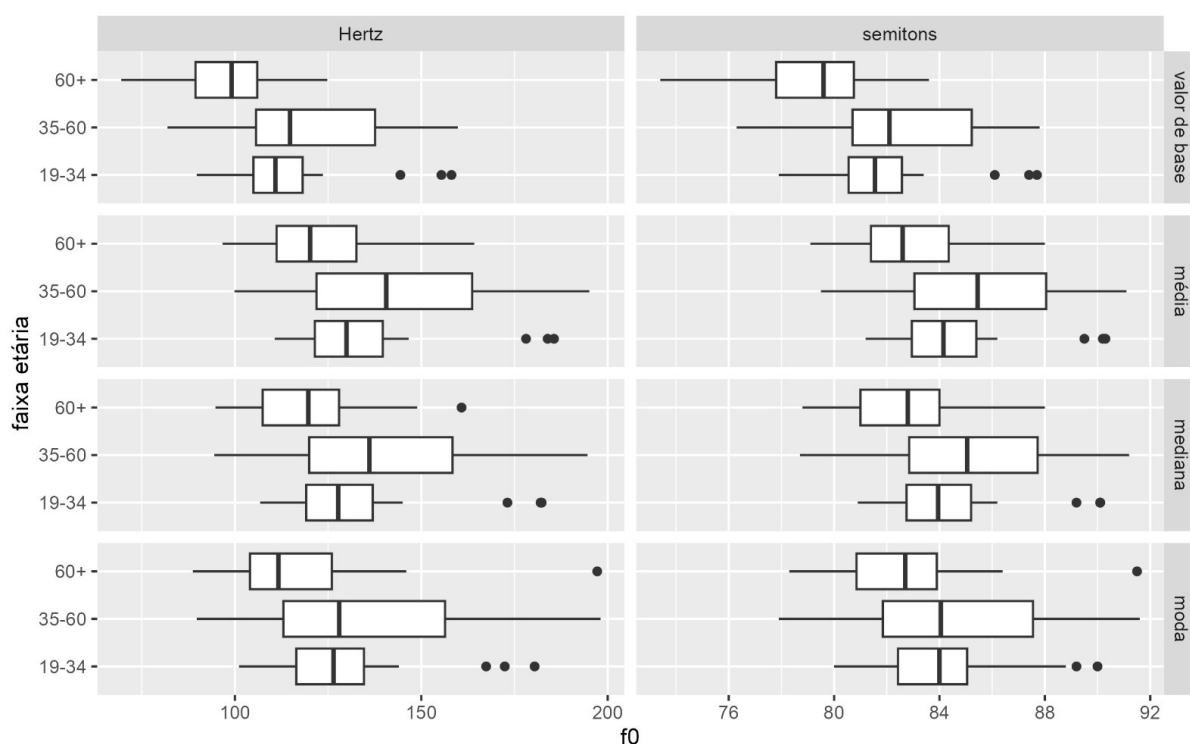


Figura 7: Diagramas de caixa dos descritores de tendência central f0 em função da escala de mensuração e das faixas etárias dos participantes.

Os valores apresentados na tabela revelam diferenças consistentes no comportamento da f0 entre os três grupos de faixas etária. As medidas de tendência central apontam para um valor de f0 mais baixo no grupo 60+ anos, com a média: 122 Hz (60+); 143 Hz (35-60); 135 Hz (19-34). Moda: 117 Hz (60+); 137 Hz (35-60); 130 Hz (19-34). Mediana: 120 Hz (60+); 140 Hz (35-60), 133 Hz (19-34).

Isso mostra que os falantes acima de 60 anos apresentam valores médios consistentemente mais baixos em todos os descritores estatísticos, em que a maior diferença se evidencia em relação ao grupo de 35 a 60 anos. Já os grupos de 19–34 e 35–60 anos mantêm valores próximos entre si em todos os estimadores, sem variações relevantes.

O padrão observado pode refletir uma associação entre a redução dos valores de  $f_0$  e fatores fisiológicos relacionados ao avanço da idade, especialmente após os 60 anos. No entanto, essa hipótese deve ser interpretada com cautela, pois o resultado diverge do que é amplamente reportado na literatura, a qual aponta tendência de aumento dos valores de  $f_0$  com o avanço da idade (VALENTE et al., 2021). Dessa forma, não se pode descartar que o efeito identificado decorra também de outros fatores, como características amostrais ou metodológicas.

Os indicadores de Desvio-padrão e Coeficiente de variação apontam um aumento de  $f_0$  com o avanço da idade. Os valores do desvio padrão são: 16,4 Hz (19–34); 19,4 Hz (35-60); 19,8 Hz (60+). Os valores do Coeficiente de variação também seguem a mesma tendência de aumento em relação ao avanço da idade, com valores de: 12,1 Hz (19–34); 13,7 Hz (35-60); 16,1 Hz (60+).

Por fim, as medidas de forma de distribuição apresentaram assimetria positiva em todas as faixas etárias, a Faixa 19-34 anos apresentou valor ligeiramente mais alto, com 0,77 Hz. Enquanto os valores da Curtose, os três grupos de faixa-etárias apresentaram valores acima de 3, o que aponta para distribuições mais concentradas ao redor da média.

Os testes estatísticos confirmam essa tendência: para a moda (Hz), com o teste de *Fligner-Killeen* foi encontrado  $p = 0,1256$ , mostrando que não há evidência estatística de que a dispersão dos dados sejam diferentes entre as faixas etárias. A aplicação do teste *ANOVA* resultou em  $p = 0,02107$ , expondo uma diferença estatisticamente significativa da Moda entre, pelo menos, duas das faixas etárias. O resultado de teste *post-hoc Wilcoxon* apresentou uma diferença significativa entre as faixas etárias 35-60 e 60+, com  $p = 0,02$ .

Para a mediana (Hz), o teste de *Fligner-Killeen* encontrou o resultado  $p = 0,03003$ , mostrando uma diferença estatisticamente significativa na dispersão. Foi aplicado o teste de *Kruskal-Wallis*, que indicou diferenças significativas estatisticamente, com  $p = 0,01101$ . O resultado de teste *post-hoc Wilcoxon* apresentou a Faixa 60+ com média de  $f_0$  mais baixa quando comparada as outras duas faixas etárias.

Os resultados encontrados para a média (Hz), o teste de *Fligner-Killeen* apontou  $p = 0,0499$ , resultado que mostra diferença estatisticamente significativa na dispersão, a aplicação do *Kruskal-Wallis* confirmou essas diferenças significativas:  $p = 0,008848$  e teste *post-hoc Wilcoxon* apontou:  $p = 0,021$  na comparação entre as faixas etárias 35-60 e 60+, e,  $p = 0,027$  na comparação entre as faixas etárias 19-34 e 60+, mais uma vez mostrando que a faixa etária 60+ apresenta valores de  $f_0$  mais baixa que as demais faixas.

Por último, os resultados encontrados para o Valor de base (Hz), o teste de *Fligner-Killeen* apontou  $p = 0,06147$ , a aplicação de *ANOVA* resultou em  $p = 8.63e-05$ , que significa uma diferença significativa estatisticamente. A aplicação de teste *post-hoc Wilcoxon* apontou:  $p = 0,00094$  na comparação entre as faixas etárias 19-34 e 60+ e,  $p = 0,00019$  na comparação entre

as faixas etárias 35-60 e 60+, mostrando diferenças significativas para a faixa etária 60+, com média mais baixa do que as demais.

Em conjunto, esses resultados demonstram que a queda dos parâmetros vocais é estatisticamente significativa e sistemática a partir dos 60 anos, enquanto não há distinção relevante entre jovens e adultos de meia-idade. A diminuição dos valores de  $f_0$  na população de idade mais avançada mostra uma tendência contrária em relação ao estudo de Valente et al. (2021), que mostra aumento do valor de  $f_0$  na população de idade mais avançada.

O estudo citado fez uma análise de valores de  $f_0$  entre as faixas-etárias de 51, 74 e 82 anos de apenas um falante, enquanto o presente trabalho realizou uma análise mais ampla, abrangendo falantes com idades desde 19 anos até 60+ anos, o que pode ser um indicador do motivo do resultado gerado chamar atenção, pelo fato de, falantes mais velhos apresentarem diminuição dos valores de  $f_0$  quando comparado com valores de falantes mais jovens.

Ainda assim, considerando o estudo de Valente et al. (2021), os resultados mostram uma certa queda do valor médio de  $f_0$  entre a faixa-etária de 74 e 82 anos, relatando uma queda significativa e testes post-hoc relatando significância estatística entre as faixas-etárias de 51 anos e 82 anos.

Por fim, o método utilizado pelo estudo se tratou da coleta de amostras através de fala espontânea de apenas uma pessoa, não possuindo uma variação de falantes, desta maneira, o resultado encontrado pelo estudo pode refletir apenas uma característica pessoal deste falante selecionado, além de que, como foi mencionado na seção 1.3, a  $f_0$  pode sofrer influência de fatores não físicos, o que pode ser a hipótese do motivo pelo qual este estudo apresentou um aumento da  $f_0$  entre a idade 51 e 74 e, depois uma certa queda entre a idade 74 e 82.

Outros resultados, também podem ser encontrados em estudos realizados anteriormente, como o de Vipperla, Renals e Frankel (2010), que aborda os efeitos da mudança de parâmetros vocálicos em reconhecimento automático de voz, em que os valores de  $f_0$  encontrados foram um pouco maior para falantes de idade mais avançada, quando comparados com falantes de meia idade. Outro estudo que apresentou aumento dos valores de  $f_0$  com o avanço da idade foi o de Taylor et al. (2020), que buscava investigar mudanças na fala e na voz de acordo com a idade.

## 5 CONSIDERAÇÕES FINAIS

De forma a servir de base para a fonética forense brasileira, por meio da geração de valores estatísticos robustos da frequência fundamental ( $f_0$ ) de falantes masculinos da cidade de São Paulo, o trabalho atingiu seu objetivo principal.

Os resultados estatísticos que foram gerados a partir da análise da  $f_0$  da população da cidade de São Paulo têm um papel importante dentro do cenário da fonética, pelo motivo de ainda não haver estudos que apresentem dados robustos de marcadores de tendência central focados somente no português falado na capital paulista. Estes resultados também possuem um grande potencial de aplicação prática, servindo como um banco de dados estatístico robusto da  $f_0$  da população paulistana e de auxílio para peritos no exame de comparação de locutores.

A obtenção de valores estatísticos da  $f_0$  nessa população é fundamental não apenas para compor um panorama acústico representativo do português falado na maior cidade brasileira, mas também para servir de referência em diversas áreas aplicadas da linguística. Os resultados gerados podem servir como base para estudos voltados à dialetologia, área que estuda as variações da língua, baseadas principalmente na distribuição geográfica entre diferentes grupos de falantes. Considerando a diversidade cultural e regional presentes na cidade de São Paulo, futuros estudos podem chegar a melhores explicações sobre a dinâmica de distribuição populacional.

Para que isso fosse possível, foram utilizadas as amostras de fala presentes no acervo Projeto SP2010 e, através dessas gravações, foram feitas estratificações por estilo de elocução e correções de valores espúrios, após esse processo, através do manuseio de scripts especializados, foram gerados os valores de  $f_0$  nos descritores estatísticos estabelecidos.

Os resultados mostraram que a distribuição da  $f_0$  varia de acordo com a região de origem dos falantes. Enquanto a Zona Sul e a Zona Central apresentaram padrões mais estáveis e homogêneos, a Zona Norte destacou-se pela maior dispersão dos valores, com significância estatística na comparação com a Zona Central. A média de  $f_0$  da Zona Central foi de 121 Hz, enquanto a da Zona Norte chegou a 139 Hz, revelando diferenças sistemáticas também em relação à Zona Oeste. Esses resultados reforçam a relevância de fatores sociofonéticos entre os falantes dessas zonas geográficas da cidade, podendo servir de ajuda para aplicações periciais.

Outro ponto relevante foi a influência da idade. Os resultados demonstraram que a queda dos parâmetros vocais é estatisticamente significativa e sistemática a partir dos 60 anos, indicando que o envelhecimento têm influência sobre a  $f_0$ . Em contrapartida, não foram observadas diferenças significativas entre falantes jovens (19–34 anos) e de meia-idade (35–60 anos), sugerindo relativa

estabilidade dos parâmetros vocais nesses dois grupos.

Em conjunto, os resultados obtidos constituem uma base sólida para uso da comunidade forense brasileira, ao oferecer parâmetros estatísticos que podem auxiliar na identificação e comparação de falantes em contextos periciais. Por conta disso, esse trabalho pode servir de base para a realização de futuras pesquisas que podem expandir esse banco de dados, incluindo até mesmo falantes de outras regiões do país. Dessa forma, será possível consolidar um referencial estatístico mais abrangente, capaz de fortalecer a prática forense nacional.

## REFERÊNCIAS BIBLIOGRÁFICAS

- ARANTES, P. Estimativas de longo termo da frequência fundamental: implicações para a fonética forense. *ReVel*, v. 12, n. 23, 2014.
- ARANTES, P. *better\_f0: A Praat script for better f0 extraction (v1.0)*. 2019. Zenodo.
- ARANTES, P. *parantes/f0-outliers: Version 1.0 (v1.0)*. 2021. Zenodo.
- BECK, J. M. Vocal profile analysis scheme: A user's manual. *Queen Margaret University College – QMUC, Speech Science Research Centre*, 2007.
- BOERSMA, P. Praat, a system for doing phonetics by computer. *Glott International*, v. 5, n. 9/10, p. 341–347, 2001.
- CARDOSO, S. A. Dialetologia. In: JUNIOR, C. F.; MOLLICA, M. C. (Ed.). *Sociolinguística, sociolinguísticas: uma introdução*. São Paulo: Contexto, 2016. p. 13–22. E-book. Disponível em: <<https://plataforma.bvirtual.com.br>>. Acesso em: 09 out. 2025.
- CUNHA, M. S. d. *Estatísticas populacionais da frequência fundamental do português brasileiro para uso em fonética forense*. Dissertação (Dissertação (Mestrado em Linguística)) — Universidade Federal de São Carlos, São Carlos, SP, 2023.
- GOLD, E.; FRENCH, P. International practices in forensic speaker comparison. *International Journal of Speech, Language and the Law*, v. 18, n. 2, p. 293–307, nov. 2011.
- HUGHES, V. S. *The definition of the relevant population and the collection of data for likelihood ratio-based forensic voice comparison*. 334 p. Tese (Tese (Doutorado em Language and Linguistic Science)) — University of York, [S. l.], 2014.
- IBGE. Relatório técnico, *Censo Demográfico 2022: Resultados preliminares*. Rio de Janeiro: [s.n.], 2022. Disponível em: <<https://censo2022.ibge.gov.br/panorama/>>. Acesso em: 21 out. 2025.
- JESSEN, M. Forensic phonetics. *Language and Linguistic Compass*, v. 2, n. 4, p. 671–711, jul. 2008.
- MENDES, R. B. Sp2010 - construção de uma amostra da fala paulistana. Projeto de pesquisa – Universidade de São Paulo, São Paulo, 2011. Projeto financiado pela Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP). 2011.
- MENDES, R. B. *Projeto SP2010: Amostra da fala paulistana*. 2013. Disponível em: <<http://projetosp2010.fflch.usp.br>>. Acesso em: 15 abr. 2024.
- MORRISON, G. Forensic voice comparison. *Expert Evidence*, v. 40, p. 13–30, sep. 2010.

NOLAN, F. *The phonetic Bases of speaker recognition*. 1. ed. Cambridge: Cambridge University Press, 1983. 240 p.

PASSETTI, R. R.; MADUREIRA, S.; BARBOSA, P. A. Voice perception on a voice messaging app: implications for forensic phonetics. In: *Proceedings of Speech Prosody 2022*. Lisboa: [s.n.], 2022. p. 490–494.

R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria, 2024. Disponível em: <<https://www.R-project.org/>>.

SILVA, G. A. Proposta de construção de um banco de dados de amostras de fala para uso forense em um arcabouço bayesiano. *Revista Brasileira de Criminalística*, v. 5, n. 1, p. 35–45, abr. 2016.

TAYLOR, S. et al. Age-related changes in speech and voice: Spectral and cepstral measures. *Journal of Speech, Language, and Hearing Research*, v. 63, n. 3, p. 647–660, 2020.

VALENTE, A. R. et al. Prosodic changes with age: A longitudinal study on a famous european portuguese native speaker. In: KARPOV, A.; POTAPOVA, R. (Ed.). *Speech and Computer*. [S.l.]: Springer International Publishing, 2021. p. 726–736. ISBN 978-3-030-87802-3.

VIPPERLA, R.; RENALS, S.; FRANKEL, J. Ageing voices: The effect of changes in voice parameters on asr performance. *EURASIP Journal on Audio, Speech, and Music Processing*, v. 2010, p. 1–10, 2010.

## ANEXOS

### A CÓDIGO SCRIPT SLICE MERGE

Projeto "Estatísticas de f0 da população masculina da cidade de São Paulo-SP para uso em fonética forense"

Corta arquivo de áudio a partir de segmentação marcada em TextGrid e junta as fatias em arquivo separado.

- Entrada: arquivo de som e TextGrid ambos com mesmo nome.

- Saída: arquivo de som com fatias unidas.

Pablo Arantes <pabloarantes@ufscar.br>

criado: 2024-05-22

```
form: "Selecione a camada com fatias marcadas"  
      natural Tier 1
```

```
endform
```

```
Select Sound plus TextGrid
```

```
nsl = numberOfSelected()
```

```
if nsl <> 2
```

```

        exitScript: "Selecione um Sound e um TextGrid."
else

    grid = selected("TextGrid")

    name$ = selected$("TextGrid")

    audio = selected("Sound")
endif

    Extract slices and merge them in a Sound object
    A matching TextGrid is also generated
    Extract non-empty intervals: tier, "no"

slices = selected("Sound")

Concatenate recoverably

audio merged = selected("Sound")

grid merged = selected("TextGrid")

Save merged objects

selectObject: audio merged

Save as WAV file: name$ + "_merged.wav"
selectObject: grid_merged
Save as text file: name$ + " merged.TextGrid"

Clean Objects list

removeObject: audio merged, grid merged, slices

writeInfo: "Done!"

```

## B CÓDIGO TABELA DE SIMULAÇÕES

```
library(ggplot2)
library(dplyr)

plot_lr <- function(mean_ref, sd_ref, mean_suspect,
  sd_suspect, offender_value, titulo) {

x <- seq(0, 220, length.out = 500)

ref_pdf <- dnorm(x, mean_ref, sd_ref)

suspect_pdf <- dnorm(x, mean_suspect, sd_suspect)

p_suspect <- dnorm(offender_value, mean_suspect, sd_suspect)

p_ref <- dnorm(offender_value, mean_ref, sd_ref)

LR <- p_suspect / p_ref

dados <- data.frame(x, ref_pdf, suspect_pdf)
ggplot(dados, aes(x)) +
  geom_line(aes(y = suspect_pdf, color = "modelo do suspeito"),
    size = 1) +
  geom_line(aes(y = ref_pdf, color = "modelo da população de
    referencia"), size = 1) +
  geom_vline(xintercept = offender_value, color = "green",
    linetype = "solid") +
  geom_point(aes(x = offender_value, y = p_ref), color =
    "green", size = 3) +
  geom_point(aes(x = offender_value, y = p_suspect), color =
    "green", size = 3) +
  geom_segment(aes(x = 0, xend = offender_value, y = p_ref, yend
```

```

    = p_ref),
  color = "green", linetype = "dotted") +
  geom_segment(aes(x = 0, xend = offender_value, y = p_suspect,
    yend = p_suspect),
    color = "green", linetype = "dotted") +
  annotate("text", x = 60, y = max(ref_pdf, suspect_pdf) * 0.9,
    label = paste0("LR = ", round(LR, 2)),
    size = 5, fontface = "bold") +
  labs(title = titulo, x = "f0 (Hz)", y = "Densidade de
    probabilidade", color = "") +
  theme_minimal(base_size = 13) +
  scale_color_manual(values = c("modelo do suspeito" = "blue",
    "modelo da população de referencia" = "red"))
}
plot_lr(135, 16.4, 170, 12, 130, "Situação 1: LR < 1 -
  evidencia favorece origens diferentes")
plot_lr(121, 18.4, 125, 15, 124, "Situação 2: LR ≈ 1 -
  hipóteses com força similar")
plot_lr(143, 19.4, 150, 10, 152, "Situação 3: LR > 1 -
  evidencia favorece mesma origem")

```

## C CÓDIGO SCRIPT PITCH TO TABLE

```

# Convert a Pitch object to spreadsheet format

# Created: 2015-04-21

# Pablo Arantes <pabloarantes@protonmail.com>

form Convert Pitch to list
  sentence Source
  sentence Output
endform

files = Create Strings as file list: "files", source$ +

```

```

    "*.Pitch"
nfiles = Get number of strings

for i to nfiles
    file$ = object$[files , i]
    pitch = Read from file: source$ + file$
    name$ = selected$("Pitch")
    ptier = Down to PitchTier
    tor = Down to TableOfReal: "Hertz"
    tab = To Table: "File"
    nrows = Get number of rows
    for r to nrows
        Set string value: r, "File", name$
    endfor
    Formula: "Time", "fixed$(self, 4)"
    Formula: "F0", "fixed$(self, 2)"
    Save as tab-separated file: output$ + name$ + ".txt"
    removeObject: pitch , ptier , tor , tab
endfor

removeObject: files
writeInfo: "Finished at ", date$(), "."

```