



**Universidade Estadual de Campinas**

**Bacharelado em Estatística**

**BEATRIZ DE VASCONSELOS NEVES**

**BRUNO DA COSTA PEIXOTO**

**RODRIGO FADEL STAHL**

**VICTÓRIA PINCINATO ANGELI**

**CAMPINAS**

**2017**

<b>1 Introdução</b>	<b>2</b>
<b>2 Limite do observável</b>	<b>3</b>
2.1 Análise descritiva	3
2.2 Análise Inferencial	5
2.2.1 Modelo com Âncora 1	5
2.2.2 Discussão dos resultados com Âncora 1	8
2.2.3 Modelo com Âncora 9	9
2.2.4 Discussão dos resultados com Âncora 9	11
2.3 Conclusão	12
<b>3 Preferência</b>	<b>13</b>
3.1 Análise descritiva	13
3.2 Análise Inferencial	16
3.3 Discussão dos Resultados	19
3.4 Conclusão	20
<b>4 Referências</b>	<b>21</b>

# 1 Introdução

Este projeto refere-se a tese de doutorado do pesquisador Alexandre Maiorino (FEC, Unicamp), com foco em acústica. O objeto da pesquisa é a sensação e percepção de reverberância pelo ouvido humano. Foram realizados experimentos para avaliar o limite do observável, isto é, qual a diferença mínima de reverberância entre dois sons que é perceptível pelo ouvido humano. Também foi conduzido um experimento para determinar a preferência, para diferentes níveis de reverberância.

Não existe um parâmetro físico para mensurar, com exatidão, a sensação de reverberância. Em nosso projeto utilizamos um parâmetro proposto pelo próprio pesquisador, baseado no tempo de decaimento do som em um campo difuso. As observações a serem descritas e estudadas foram ordenadas segundo este parâmetro.

Os experimentos foram realizados com voluntários, vinte no total, em dias distintos. Em uma sala com isolamento acústico e com fones de ouvidos próprios. Cada voluntário respondia a dois questionários em sequência. Tais questionários eram autoexplicativos, o pesquisador não explicava o experimento, para evitar diferenciação de tratamento entre os voluntários.

No primeiro, referente à percepção do limite do observável, um áudio era definido como âncora, sendo este o de menor reverberância. O âncora era combinado com outros nove áudios, incluindo o próprio âncora, formando nove pares. Todos os áudios tratavam-se do mesmo trecho de uma obra, diferindo entre si apenas por sua reverberância, da menor para a maior, esta diferença foi obtida digitalmente em um estúdio. Os pares eram ouvidos de forma aleatória por cada voluntário e o voluntário ouvia duas vezes cada par, totalizando dezoito pares. Após ouvir cada par, deveria-se responder à questão “Existe diferença entre os áudios?” com resposta dicotômica “Sim” ou “Não”. Então o âncora era trocado pelo áudio de maior reverberância e o procedimento repetido para outros dezoito pares.

O segundo questionário, refere-se a preferência. Foram escolhidos três áudios, um de menor reverberância, um intermediário e um de maior reverberância. Assim como no primeiro questionário este áudios são o mesmo trecho de uma obra, digitalmente diferenciados apenas por sua reverberância. Para cada voluntário o primeiro áudio era fixado, e pareado com um segundo escolhido aleatoriamente entre os dois restantes. Cada áudio era

fixado como o primeiro três vezes, portanto para cada indivíduo foram formados nove pares. Os áudios eram ouvidos aleatoriamente, após ouvir cada par o voluntário respondia a questão “Qual dos áudios é seu preferido?” com resposta dicotômica “Primeiro áudio” e “Segundo áudio”.

A partir dos experimentos descritos pretende-se alcançar dois objetivos. O primeiro é estabelecer o limite observável segundo o parâmetro proposto. Ou seja, dados os níveis de reverberância, deseja-se determinar a partir de qual deles o ouvido humano consegue diferenciar o âncora de um áudio distinto. O outro objetivo é determinar qual nível de reverberância é o preferido. Uma vez que cada objetivo está relacionado a um experimento, estes serão analisados separadamente.

## 2 Limite do observável

### 2.1 Análise descritiva

Proporções de pares (Âncora - Áudio) considerados iguais, ordenados por nível de reverberância, para as diferentes âncoras

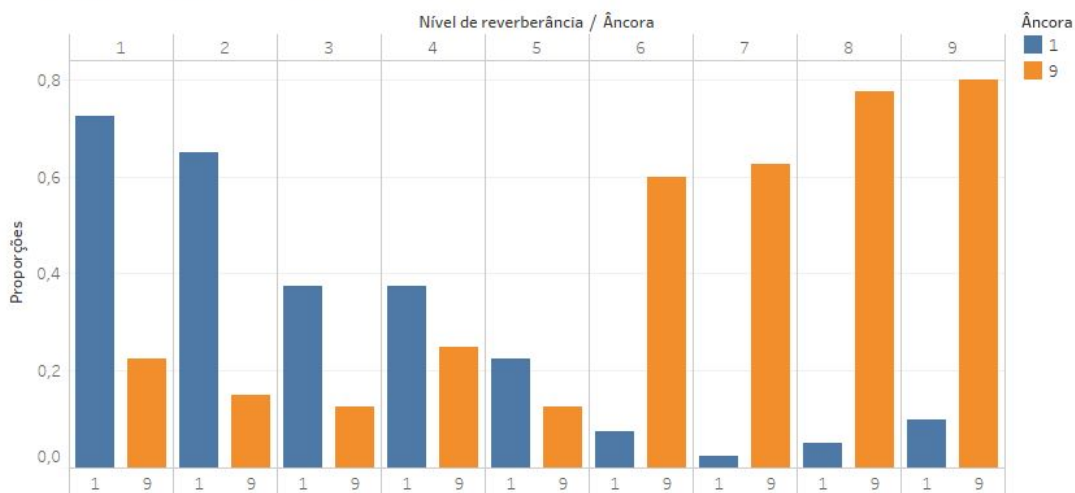


Figura 2.1: Gráfico de barras com as proporções de pares considerados iguais. Os níveis de reverberância foram ordenados de 1 a 9, do de menor valor para o maior. Em azul os áudios são comparados com o áudio 1 (A âncora é o áudio 1), analogamente em laranja os áudios são comparados com o áudio 9 (A âncora é o áudio 9).

O gráfico da figura 2.1 foi construído ordenando-se os áudios de um a nove, do de maior reverberância para o de menor. Fixado uma Âncora (em azul, áudio 1; em Laranja, áudio 9) e comparando-se um a um com todos os áudios. Para cada par (Âncora - Áudio) é mostrada a proporção de vezes em que não foi possível diferenciar o áudio da âncora.

Nota-se o decaimento das proporções conforme os áudios se distanciam da âncora. Percebe-se também que nem mesmo quando os áudios eram iguais a proporção de todas as pessoas reconhecem como sendo iguais. Quando perguntado ao pesquisador, este sugeriu que tais efeitos podem devidos a natureza do experimento.

Quando a âncora foi o áudio 1, os dois primeiros áudios (1 e 2) tem proporções próximas de 70%, sugerindo que possivelmente não são facilmente distinguíveis, já os próximos dois áudios (3 e 4) tem proporções entre 30% e 40%, o que sugere que nesta região as pessoas não têm certeza quanto a diferença, e os outros áudios (5, 6, 7, 8 e 9) tem proporções de classificação abaixo de 30%, conjectura-se que nesta região os erros de classificação vêm de outra fonte de variabilidade.

Tendo-se o áudio 9 como âncora, os áudios 8 e 9 têm proporção de classificação muito próximas (cerca de 80%), dificilmente estes áudios serão tidos como diferentes. Os áudios 6 e 7 estão em um patamar menor, mas ainda próximo (cerca de 60%), é possível que as pessoas tenham dificuldade de discernir estes áudios da âncora. Quanto ao resto dos áudios (5, 4, 3, 2 e 1), todos possuem uma proporção de classificação de igualdade abaixo de 30%, na maioria das vezes os voluntários acertaram ao diferenciá-los da âncora.

Também é possível construir um gráfico similar ao da figura 2.1 para verificar as diferenças entre os indivíduos, no entanto este é muito extenso para colocar aqui, devido ao número de indivíduos. Ele pode ser acessado através do link:

[https://public.tableau.com/profile/bruno.da.costa.peixoto#!/vizhome/JNB\\_descritiva/Planilha1](https://public.tableau.com/profile/bruno.da.costa.peixoto#!/vizhome/JNB_descritiva/Planilha1)

Percebe-se que algumas pessoas para certas observações não mantiveram a concordância, isto é, em uma questão eles afirmam que os áudios são diferentes em outra que são iguais.

## 2.2 Análise Inferencial

Para resolver o problema do limite do observável, optamos por trabalhar com as proporções em que cada indivíduo respondia que não era possível diferenciar os áudios, para cada par de áudios (âncora - áudio), a fim de estimar a proporção de vezes em que não se diferencia o âncora do respectivo áudio. Decidiu-se utilizar um modelo de regressão beta com efeito aleatório (modelo misto com distribuição beta) para modelar a proporção, devido ao fato de seu suporte estar entre 0 e 1, e cada indivíduo apresentar respostas para mais de um par de áudios gerando dependência das observações. Separou-se este problema em 2 modelos, sendo um para as respostas com o áudio mais seco definido como âncora (áudio 1) e o outro para as respostas com o áudio mais reverberante definido como âncora (áudio 9). Foi utilizado o modelo implementado usando o pacote *gamlss* no *software* R.

### 2.2.1 Modelo com Âncora 1

Ao estimar a proporção de vezes que não foi possível diferenciar o áudio  $j$  da âncora 1,  $Y_j$ , o modelo utilizado assume que cada observação  $y_{ij}$ , proporção de vezes que o indivíduo  $i$  não consegue diferenciar o áudio  $j$  da âncora 1, vem da seguinte distribuição:

$$Y_j \sim \text{Beta}(\mu_j, \phi), \text{ com } i \in \{10, 11, 12, \dots, 29\} \text{ e } j \in \{1, 2, 3, \dots, 9\}.$$

Onde o parâmetro  $\mu_j$  é a média esperada de  $Y_j$  e  $\phi$  é seu respectivo parâmetro de precisão.

Estima-se os parâmetros  $\mu_j$ , aplicando-se o algoritmo EM estimar a máxima verossimilhança da seguinte expressão:

$$g(\mu_j) = x_1 \beta_1 + x_2 \beta_2 + x_3 \beta_3 + x_4 \beta_4 + x_5 \beta_5 + x_6 \beta_6 + x_7 \beta_7 + x_8 \beta_8 + x_9 \beta_9 + \eta.$$

Onde  $g(\cdot)$  é a função de ligação logito, ou seja,  $g(\mu_j) = \log(\mu_j/(1 - \mu_j))$ . O parâmetro  $\eta$  é o efeito aleatório causado pela dependência nas observações por indivíduo  $i$ , com

$\eta_i \sim N(0, \sigma^2)$ . E  $x_k = 0 \forall k$ , exceto  $k = 1$  e  $k = j$ , se  $j \neq 1$ , que  $x_k = 1$ . Ou seja,  $x_k$  assume valor 1 para  $k$  correspondente aos áudios comparados, âncora( $k=1$ ) e áudio 2( $k=j$ ).

São estimados os parâmetros  $\beta_k$ , onde o par referente ao âncora( $k=1$ ), é fixado como intercepto na metodologia casela de referência, de modo a facilitar o alcance do objetivo. Também é estimado  $\phi$ , assumindo-se que é igual para todas as categorias.

Os resultados encontram-se na tabela abaixo:

Tabela 2.1: Estimativas dos parâmetros  $\beta_k$  no modelo ajustado para o áudio 1 fixado como âncora, com seus respectivos p-valores do teste de nulidade.

<b>Âncora 1</b>		
<b>Parametro</b>	<b>Estimativa <math>\beta</math></b>	<b>p-valor</b>
$\beta_1$ (intercepto)	0.8574	0.0022
$\beta_2$	-0.2715	0.4832
$\beta_3$	-1.3695	0.0006
$\beta_4$	-1.3787	0.0006
$\beta_5$	-1.8609	<0.0001
$\beta_6$	-2.3004	<0.0001
$\beta_7$	-2.4565	<0.0001
$\beta_8$	-2.3919	<0.0001
$\beta_9$	-2.3101	<0.0001
$\phi$	1.0745	<0.0001

A tabela 2.1 apresenta os valores estimados dos parâmetros  $\beta_k$ , e o p-valor associado a seu respectivo teste de nulidade. Ou seja, ele testa se  $\beta_1 = 0$ , e se  $\beta_k = \beta_1$ , para  $k \in \{2, 3, \dots, 9\}$  e se  $\phi = 0$ .

Com o resultado apresentado na tabela 2.1, somos capazes de estimar a proporção esperada de vezes que um indivíduo não consegue diferenciar a âncora 1 dos demais áudios ( $\mu_j$ ), utilizando anti logito em  $g(\mu_i)$ , que é calculado através dos valores estimados de  $\beta_k$ . Consequentemente, podemos realizar a predição das proporções para os nove áudios, e utilizar a metodologia de *bootstrap* para estimar o intervalo de confiança para cada predição.

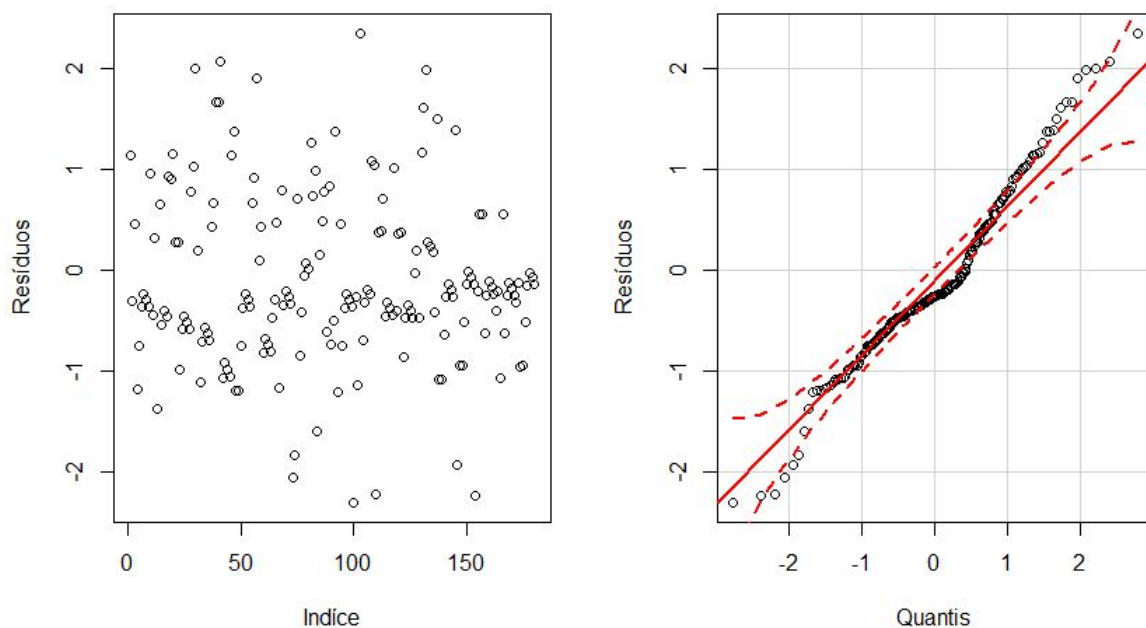
A metodologia de *bootstrap* consiste em retirar  $n$  amostras de mesmo tamanho com reposição da amostra original, nestas são estimados os parâmetros utilizando a mesma metodologia. Assim é obtida uma estimativa da densidade original para o qual são calculados os intervalos de confiança. Os resultados obtidos através desta metodologia encontram-se na tabela 2.2 a seguir:

*Tabela 2.2: Predição do modelo ajustado, fixado âncora 1 para cada par de áudio, e os respectivos intervalos de confiança e proporção amostral de vezes em que não foi possível diferenciar os áudios.*

<b>Âncora 1</b>					
<b>Áudio 1</b>	<b>Áudio 2</b>	<b>Predição</b>	<b>Amostra</b>	<b>IC</b>	
				<b>2,50%</b>	<b>97,50%</b>
1	1	0.7021	0,7250	0.5566	0.7968
1	2	0.6424	0,6500	0.5000	0.7554
1	3	0.3746	0,3750	0.2471	0.5258
1	4	0.3725	0,3750	0.2367	0.5273
1	5	0.2682	0,2250	0.1784	0.3760
1	6	0.1910	0,0750	0.1432	0.2639
1	7	0.1681	0,0250	0.1390	0.2200
1	8	0.1773	0,0500	0.1432	0.2322
1	9	0.1895	0,1000	0.1516	0.2691

Podemos notar que a predição apresenta um valor muito próximo do encontrado para os pares mais difíceis de distinguir, porém começa a afastar um pouco para os pares mais distintos entre si. Porém como o número de observações foi muito pequeno, e é natural possuir uma diferença do valor predito em relação ao valor observado, esses resultados são esperados.





*Figura 2.2 : Gráfico de resíduos ajustados x índice e gráfico de quantil-quantil com bandas de envelopes para o modelo ajustado fixado âncora 1.*

Observando o gráfico de resíduos ajustados x índice da figura 2.2 , podemos observar que a suposição de homocedasticidade parece ser razoável, pois os dados encontram-se bem distribuídos e não há indícios de nenhum padrão. O outro gráfico da figura 2. nos faz suspeitar quanto a suposição de normalidade, pois os pontos saem um pouco das bandas no centro, e as caudas parecem longas.

### 2.2.2 Discussão dos resultados com Âncora 1

O objetivo é encontrar a partir de qual áudio as pessoas conseguem diferenciar os áudios em relação ao âncora. Para isso, a estratégia adotada foi comparar as estimativas dos parâmetros referentes a cada áudio. Isto pode ser observado através da terceira coluna da tabela 2.1, pois o p-valor é referente ao teste de nulidade já mencionado.

Consideramos um nível de significância de 5%, então se o p-valor for maior de 0.05 indica que não rejeitamos a hipótese nula, , ou seja, indica que o efeito do intercepto não é significativo, ou que os efeitos dos parâmetros não possuem diferença significativa em relação ao efeito do intercepto.

Observando a tabela 2.1, notamos que o p-valor de  $\beta_2 > 0.05$ , ou seja, o incremento causado pelo áudio 2 não possui diferença significativa em relação ao incremento causado pelo áudio 1. Enquanto o p-valor do  $\beta_3 < 0.05$ , indicando diferente significativa em relação aos incrementos do áudio 3 e do áudio 1. Portanto, definimos o áudio 2 como a nossa faixa de corte sobre o limite do observável.

Uma observação a ser feita, é que a partir das análises de resíduos a suposição de normalidade não é totalmente satisfeita, indicando que existe uma estrutura aleatória não explicada. A partir dos intervalos de confiança nota-se que a borda inferior referente ao áudio 1 (âncora) se aproxima da borda superior dos intervalos referentes aos áudios 3 e 4. Isto não necessariamente implica que estes são iguais, mas deve-se tomar cuidado ao tomar decisões supondo que estes são diferentes.

### 2.2.3 Modelo com Âncora 9

O modelo ajustado, para o âncora 9, é o mesmo do modelo anterior, com o âncora fixado no áudio 1. Porém desta vez, o intercepto é o áudio 9(novo âncora).

O modelo e as estimativas dos parâmetros foram ajustadas e estimadas seguindo a mesma metodologia, e encontram-se na tabela a seguir:

*Tabela 2.3: Estimativas dos parâmetros  $\beta_k$  no modelo ajustado para o áudio 9 fixado como âncora, com seus respectivos p-valores do teste de nulidade.*

<b>Âncora 9</b>		
<b>Parametro</b>	<b>Estimativa</b>	<b>p-valor</b>
$\beta_1$	-2.0507	<0.0001
$\beta_2$	-2.2777	<0.0001
$\beta_3$	-2.3273	<0.0001
$\beta_4$	-1.9638	<0.0001
$\beta_5$	-2.3377	<0.0001
$\beta_6$	-0.7008	0.0750
$\beta_7$	-0.5616	0.1567
$\beta_8$	-0.0987	0.7991
$\beta_9$ (intercepto)	1.0660	0.0002
$\phi$	1.1773	<0.0001

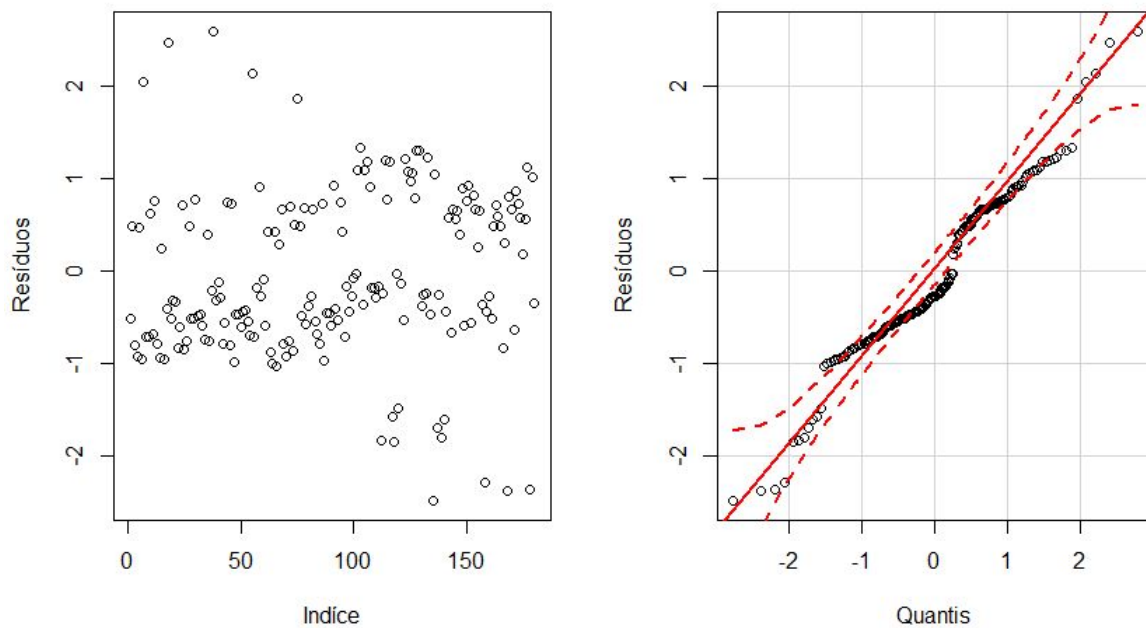
A tabela 2.3 apresenta os valores estimados dos parâmetros  $\beta_k$ , e o p-valor associado a seu respectivo teste de nulidade. Ou seja, ele testa se  $\beta_9 = 0$ , e se  $\beta_k = \beta_9$ , para  $k \in \{1, 2, 3, \dots, 8\}$  e se  $\phi = 0$ .

Da mesma forma que foi realizado no modelo anterior, com esses resultados somos capazes de estimar a proporção esperada de vezes que um indivíduo não consegue diferenciar a âncora 9 dos demais áudios ( $\mu_j$ ), utilizando anti logito em  $g(\mu_i)$ , que é calculado através dos valores estimados de  $\beta_k$ .

Consequentemente, podemos realizar a predição das proporções para os nove áudios, e utilizar a metodologia de *bootstrap* para estimar o intervalo de confiança para cada predição. Estes valores são encontrados na tabela a seguir:

*Tabela 2.4: Predição do modelo ajustado, fixado âncora 9 para cada par de áudio, e os respectivos intervalos de confiança encontrados via Bootstrap e proporção amostral de vezes em que não foi possível diferenciar os áudios.*

Âncora 9					
Áudio 1	Áudio 2	Predição	Amostra	IC	
				2,50%	97,50%
9	1	0.2719	0,2250	0.1782	0.3920
9	2	0.2294	0,1500	0.1576	0.3276
9	3	0.2207	0,1250	0.1481	0.3032
9	4	0.2894	0,2500	0.1955	0.4076
9	5	0.2189	0,1250	0.1670	0.2742
9	6	0.5903	0,6000	0.4333	0.7425
9	7	0.6234	0,6250	0.4346	0.7658
9	8	0.7245	0,7750	0.6149	0.8150
9	9	0.7438	0,8000	0.6255	0.8283



*Figura 2.3 : Gráfico de resíduos ajustados x índice e gráfico de quantil-quantil com bandas de envelopes para o modelo ajustado fixado âncora 9.*

No gráfico Resíduos x índice, podemos observar uma tendência decrescente. Este é um indício que a suposição de homocedasticidade não está satisfeita. Através do outro gráfico da figura 2., podemos notar que a suposição de normalidade não parece razoável, uma vez que os pontos apresentam comportamento anormal e fogem das bandas em alguns intervalos.

## 2.2.4 Discussão dos resultados com Âncora 9

Da mesma forma que no modelo anterior, utilizaremos o teste de nulidade para decidir a faixa de corte. Podemos observar que o p-valor de  $\beta_6$ ,  $\beta_7$  e  $\beta_8$  são maiores que 0.05, portanto não há diferença significativa em relação ao incremento causado pelo áudio 9 e pelos áudios 6, 7 e 8. Enquanto o p-valor do  $\beta_5 < 0.05$ , indicando que os incrementos são diferentes significativamente. Portanto, o áudio 6 é uma boa faixa de corte para o limite do observável.

Quanto aos resíduos, estes fogem da normalidade. Apontando que existe algum efeito aleatório não explicado. No entanto ao se observar os intervalos de confiança, nota-se que a borda inferior referente ao áudio 9 (o âncora) está distante da borda superior dos áudios em

que se diferenciam do âncora. Assim, mesmo que o modelo não capte totalmente a estrutura de erros conjectura-se que os resultados são válidos.

## 2.3 Conclusão

Conclui-se que quando o âncora é o áudio 1 não é possível diferenciar do áudio 2, mas se diferencia dos demais. Quando o âncora é o áudio 9 não é possível diferenciar dos áudio 8, 7 e 6, mas se diferencia dos demais. É válido ressaltar que o modelo não teve performance ótima, mas acredita-se que são resultados satisfatórios

## 3 Preferência

### 3.1 Análise descritiva

Para facilitar a interpretação dos resultados foi definida uma nomenclatura para cada um dos áudios. Desta forma o áudio mais reverberante foi chamado de “Áudio 1”, o intermediário foi definido como “Áudio 2” o de menor reverberância foi definido como “Áudio 3”. Quando se refere aos pares, o “Par 1-2” é o par formado pelos áudios “1” e “2”. Se suposto que a ordem dos áudios importa, o “Par 1-2” é aquele em que o primeiro é o “Áudio 1” e o segundo o “Áudio 2”, o “Par 2-1” é aquele em que o primeiro é o “Áudio 2” e o segundo o “Áudio 1”. Essa nomenclatura é estendida a todas as combinações de áudios.

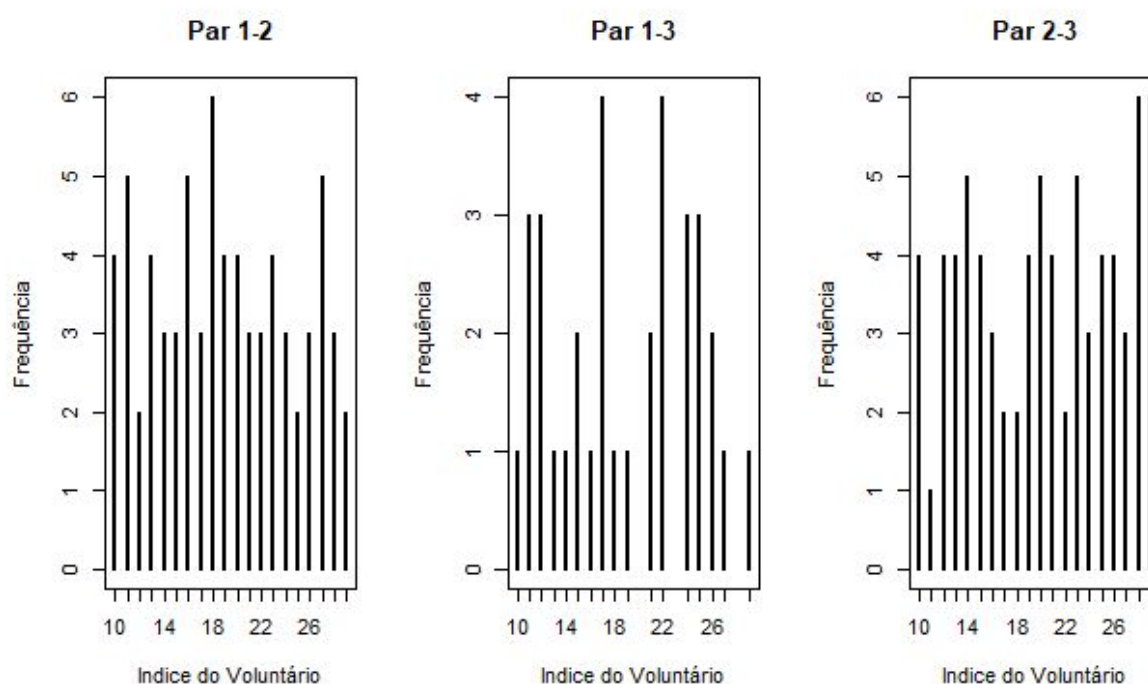
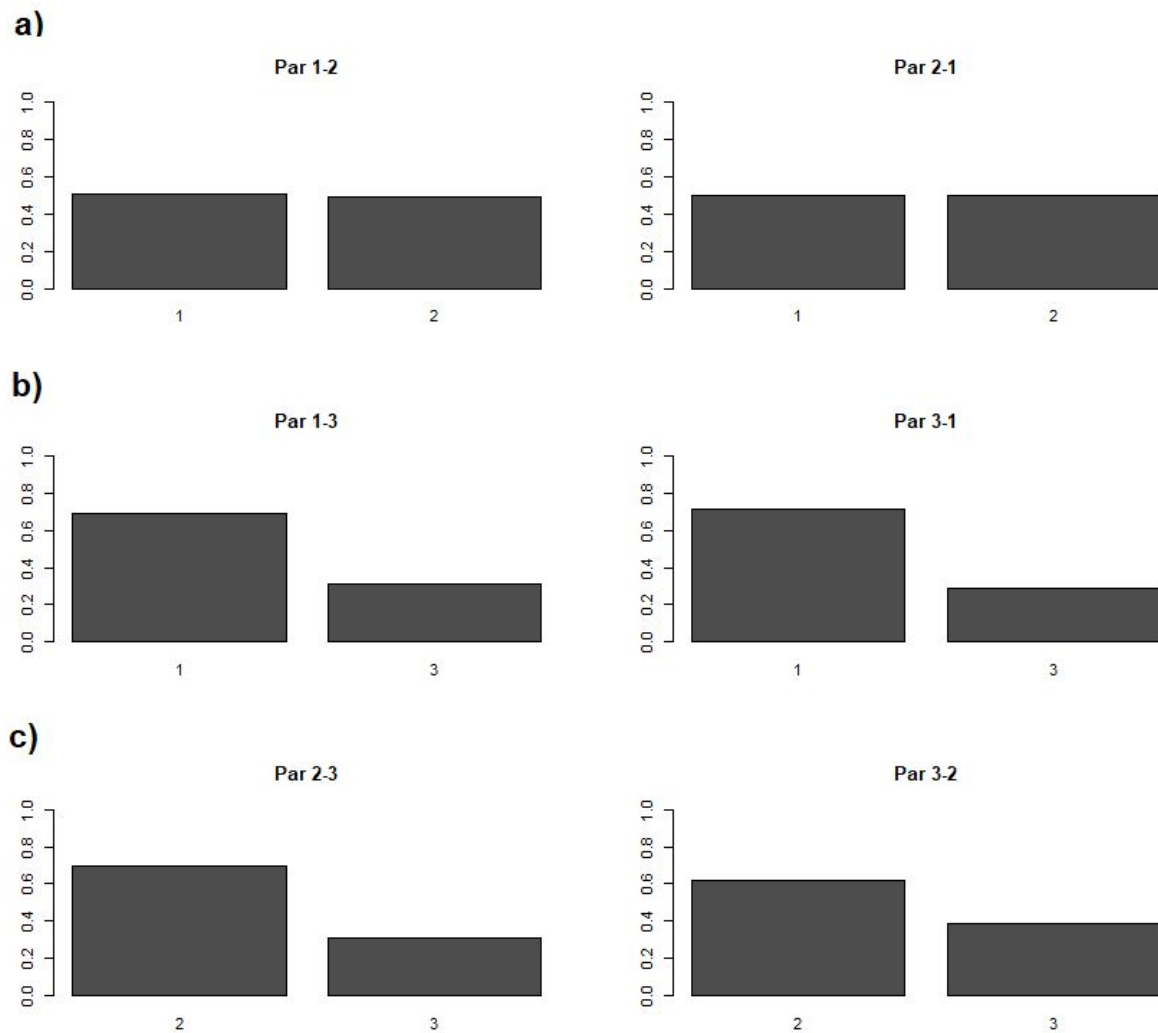


Figura 3.1: Gráfico de Barras com a frequência com que os pares foram sorteados para cada voluntário.

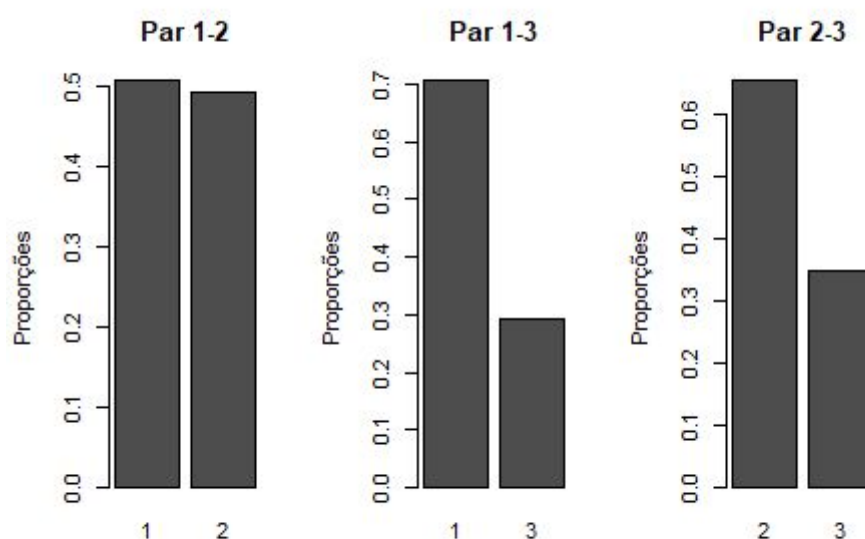
A partir da figura 3.1, é possível notar que os pares não foram igualmente distribuídos entre os indivíduos, sendo que para alguns indivíduos não foi sorteado nenhum par 1-3. Também percebe-se que existe menor quantidade de pares 1-3 do que os demais pares.

De fato foram sorteados no total 71 pares 1-2, 34 pares 1-3 e 75 pares 2-3, ignorando-se a ordem dos pares. Conclui-se que, devido a forma como os experimentos foram conduzidos, a amostra não está balanceada.



*Figura 3.2 : Gráficos de Barra comparando a proporção de preferência quando a ordem dos áudios é trocada. O primeiro gráfico da figura 3.2 a. mostra a proporção de vezes que os áudios 1 e 2 foram escolhidos como prediletos, quando o primeiro era o áudio 1, o segundo mostra a proporção de vezes que os áudios 1 e 2 foram escolhidos como prediletos, quando o primeiro era o áudio 2. O raciocínio é estendido de forma análoga para as figuras 3.2 b .e 3.2 c.*

Os gráficos da figura 3.2, foram construídos para verificar se a ordem em que os áudios estavam nos pares influencia na preferência dos indivíduos. Nas três figuras, as proporções das preferências para cada par manteve-se praticamente a mesma quando a ordem dos áudios era trocada, mesmo para os pares 2-3 e 3-2, em que a diferença é mais perceptível, não passa da casa dos 10%. Portanto há indícios de que a ordem dos áudios não interfere na escolha dos voluntários.



*Figura 3.3: Gráficos de Barra com as proporções de predileção para cada par, o primeiro gráfico mostra o resultado quando são comparados os áudios 1 e 2, o segundo os áudios 1 e 3 e o terceiro os áudios 2 e 3. A partir destes gráficos é possível comparar os pares, já que a proporção desconsidera o tamanho da amostra.*

Uma vez que o número de pares é distinto para cada combinação de áudios, a estratégia escolhida para comparar a preferência pelos áudios foi usar a proporção de predileção entre os pares. A partir da Figura 3.3 nota-se que quando comparados os áudios 1 e 2, as proporções de predileção pouco diferem. No entanto os áudios 1 ou 2 se saem melhor quando comparados com o áudio 3. O que se pode supor a partir destes gráficos é que não existe diferença de predileção entre os áudios 1 e 2, mas estes são preferidos quando comparados ao áudio 3.



## 3.2 Análise Inferencial

Para verificar se de fato a ordem dos áudios importa, será aplicado um teste de homogeneidade para duas populações, tal como descrito por Azevedo(2017). As observações serão dispostas em tabelas de contingência ( Tabelas 3.1, 3.2 e 3.3). Assumindo que as tabelas são geradas pelo produto de binomiais independentes, segue a descrição do teste.

*Tabela 3.1: Tabela de contingência dos pares 1-2 e 2-1, com as preferências observadas em cada par e sua respectiva proporção estimada (Estimador de máxima verossimilhança)*

	Prefere áudio 1		Prefere áudio 2		Total
	Observados	Proporção	Observados	Proporção	
Par 1-2	24	~51%	23	~49%	47
Par 2-1	12	50%	12	50%	24
Total	36	~51%	35	~49%	71

*Tabela 3.2: Tabela de contingência dos pares 1-3 e 3-1, com as preferências observadas em cada par e sua respectiva proporção estimada (Estimador de máxima verossimilhança)*

	Prefere áudio 1		Prefere áudio 3		Total
	Observados	Proporção	Observados	Proporção	
Par 1-3	9	~69%	5	~31%	13
Par 3-1	15	~71%	6	~29%	21
Total	24	~71%	11	~29%	35

*Tabela 3.3: Tabela de contingência dos pares 2-3 e 3-2, com as preferências observadas em cada par e sua respectiva proporção estimada (Estimador de máxima verossimilhança)*

	Prefere áudio 2		Prefere áudio 3		Total
	Observados	Proporção	Observados	Proporção	
Par 2-3	25	~69%	11	~31%	36
Par 3-2	24	~62%	15	~38%	39
Total	49	~65%	26	~35%	75

As hipóteses de interesse são da forma:

$H_0$ : As proporções independem da ordem dos pares

$H_1$ : O contrário acontece

Matematicamente podem ser expressas como:

$H_0$ :  $\theta_{11} = \theta_{21}$

$H_1$ :  $\theta_{11} \neq \theta_{21}$

Onde  $\theta_{il}$  é o parâmetro de proporção associado às caselas das primeiras colunas das tabelas de contingência, note que se  $\theta_{11} = \theta_{21}$ , então  $\theta_{12} = \theta_{22}$ .

A estatística do teste é dada por:

$$Z_t = \frac{\hat{\theta}_{11} - \hat{\theta}_{21}}{\sqrt{\hat{\theta}(1-\hat{\theta})\left(\frac{1}{n_{1.}} + \frac{1}{n_{2.}}\right)}}$$

Onde  $n_{1.}$  e  $n_{2.}$  são o número de observações em cada par,  $\hat{\theta}_{il}$  é o estimador de máxima verossimilhança de  $\theta_{il}$  e  $\hat{\theta} = \frac{n_{1.}\hat{\theta}_{11} + n_{2.}\hat{\theta}_{21}}{n_{1.} + n_{2.}}$

Sob  $H_0$   $Z_t \approx N(0,1)$ , isto é,  $Z_t$  se aproxima de uma distribuição normal padrão quando satisfeita a hipótese nula. Quanto maior  $|Z_t|$ , mais evidência contra a hipótese nula.

*Tabela 3.4: Resultado dos testes de hipóteses para os pares*

Pares	$Z_t$	p-valor
1-2 e 2-1	0,085	0.932
1-3 e 3-1	0,137	0,891
2-3 e 3-2	0,179	0,472

Os resultados do teste são exibidos na tabela 3.4. O teste foi implementado utilizando-se o pacote R. Fixando um nível de significância de 5%, Não se rejeita a hipótese nula para nenhum dos pares. Portanto estatisticamente pode-se afirmar que não importa a ordem dos áudios. Portanto a análise seguirá desconsiderando-se a ordem.

Para testar a igualdade das proporções será usado o teste assintótico com correção de continuidade. Assumindo que as observações vem de uma de uma distribuição binomial, este teste aproxima a distribuição por uma normal padrão

Sejam as hipóteses:

$H_0$ : As proporções são iguais

$H_1$ : As proporções são diferentes

Matematicamente:

$H_0$ :  $\theta = 0,5$

$H_1$ :  $\theta \neq 0,5$

Será usada a seguinte estatística do teste:

$Q = Z_c^2$ , onde  $Z_c$  é dado por:

$$Z_c = \left\{ \frac{(\hat{p} - p_0) + 1/(2n)}{\sqrt{p_0(1-p_0)/n}} \text{ se } \hat{p} - p_0 < 0 ; \frac{(\hat{p} - p_0) - 1/(2n)}{\sqrt{p_0(1-p_0)/n}} \text{ se } \hat{p} - p_0 > 0 \right\}$$

Onde,  $\hat{p}$  é a estimativa da proporção,  $p_0 = 0,5$  e  $n$  é o tamanho da amostra. Sendo  $Z_c \approx N(0,1)$ , então  $Q \approx X^2_1$ , i.e., a estatística  $Q$  se aproxima de uma estatística Chi-quadrado com um grau de liberdade.

*Tabela 3.5: Resultados dos testes de hipóteses para cada par, contendo seus respectivos valores da estatística  $Q$ , p-valor, a proporção de preferência estimada para o primeiro áudio do par e seu intervalo de 95% confiança assintótico.*

Par	$X^2_1$	p-valor	Proporção Estimada	IC(95%)
1-2	$\sim 0$	1	0.51	[0,39 ; 0,63]
1-3	4,97	0,03	0,71	[0,52 ; 0,84]
2-3	6,45	0,01	0,65	[0,53 ; 0,76]

Os resultados dos testes, tabela 3.5, foram obtidos utilizando-se a função *prop.test* implementada no pacote R. Sob um nível de significância de 5%, pode-se afirmar que não há diferença significativa entre as proporções no par 1-2, mas que existe evidência suficiente para rejeitar a hipótese nula em favor da alternativa, para os pares 1-3 e 2-3, ou seja eles são estatisticamente diferentes. A partir das proporções estimadas e seus intervalos de confiança é possível identificar a preferência dentro dos pares.

### 3.3 Discussão dos Resultados

Primeiro ressalta-se que as análises estatísticas supõem que as amostras venham de distribuições binomiais, o que não é verdade. Pela forma que procedeu-se o experimento, onde o mesmo par era avaliado mais de uma vez por cada indivíduo, sendo assim as observações não vem de bernoullis independentes. Ao não considerar este fato, deixa-se de levar em conta um efeito aleatório, que pode implicar na subestimação do erro.

Outro fato a ser considerado é o não balanceamento dos dados. Como aponta a figura 3.1, na análise descritiva, o número de pares não é homogêneo para todos indivíduos. Isto dificulta apontar resultados mais precisos e pode adicionar erro ao resultado

Levando em conta o que é apontado, deve-se considerar o resultados dos testes com cuidado. Nos testes de homogeneidade das amostras, tabela 4, os p-valores estão distantes da região de rejeição, possivelmente, com um teste mais adequado, a hipótese nula não seria rejeitada. Já nos testes de proporção, tabela 5, somente para o par 1-2 o p-valor está distante da borda da região de rejeição, para os pares 1-3 e 2-3 os p-valores 0,03 e 0,01 estão próximos do nível de significância 0,05, desta forma, deve-se ficar atento ao tomar decisões baseadas nestes testes.

### 3.4 Conclusão

A partir dos resultados conclui-se que não há diferença significativa entre os áudios 1 e 2. Quando comparado aos outros dois, há uma predileção menor ao áudio 3. Vale ressaltar que estes resultados não devem ter caráter probatório, mas indicial.

## 4 Referências

Wilson, E.B. (1927) Probable inference, the law of succession, and statistical inference. J. Am. Stat. Assoc., 22, 209–212.

Newcombe R.G. (1998) Two-Sided Confidence Intervals for the Single Proportion: Comparison of Seven Methods. Statistics in Medicine 17, 857–872.

Newcombe R.G. (1998) Interval Estimation for the Difference Between Independent Proportions: Comparison of Eleven Methods. Statistics in Medicine 17, 873–890.

Rigby R.A. and Stasinopoulos D.M. (2005). Generalized additive models for location, scale and shape,(with discussion), Appl. Statist., 54, part 3, pp 507-554.

Azevedo, C. L. N.(2017), Análise de tabelas de contingência, Parte 1. Notas de Aula no link [http://www.ime.unicamp.br/~cnaber/aula\\_tabelas\\_de\\_contingencia\\_ADD\\_1S\\_2017.pdf](http://www.ime.unicamp.br/~cnaber/aula_tabelas_de_contingencia_ADD_1S_2017.pdf)