

# Relatório Bruna

## Bibliotecas

## Saneamento dos dados

## Determinação do tamanho da amostra

Para determinar o tamanho da mostra usaremos a seguinte metodologia: primeiro, calcularemos o tamanho total da amostra, sem considerar os estratos, por meio do plano AAS e, em seguida, usaremos o plano amostral AEpr para determinar o tamanho amostral de cada estrato. Convém ressaltar que o referido AAS é subdividido em outras duas subcategorias: a primeira sendo amostragem aleatórias simples com reposição (AASc) e a segunda amostragem aleatória simples sem reposição (AASs). No primeiro caso, temos que a cada unidade amostral aleatoriamente selecionada, ela é reposta na população podendo ser selecionada novamente nas etapas seguintes. Tal procedimento garante que cada unidade amostral coletada seja **independente**. No plano AASs, por outro lado, ao sortearmos um elemento da população, este não poderá ser sorteado novamente nas etapas seguintes, de modo que, as retiradas não são independentes pois a probabilidade de sortear o elemento seguinte é alterada pelo elemento sorteado anteriormente. De modo geral, a escolha por um procedimento ou outro fica a critério do pesquisador. O plano AASc apresenta a importante propriedade da **independência** o que pode facilitar enormemente a realização de inferências futuras. Por outro lado, o efeito do planejamento (EPA), medida usada para quantificar a eficácia de determinado plano amostra, da AASs é sempre melhor que o AASc, com exceção do caso  $n = 1$ . Desse modo, cabe ao pesquisador pesar as vantagens e desvantagens, para fazer sua escolha. Abaixo, escrevemos a definição do EPA

$$EPA = \frac{Var_{AASs}[\bar{y}]}{Var_{AASc}[\bar{y}]} = \frac{\frac{(1-f)S^2}{n}}{\frac{\sigma^2}{n}} = \frac{N-n}{N-1}$$

onde notamos que  $EPA \leq 1$  sempre, de modo que o numerador é menor que o denominador, implicando que o plano AASs sempre é mais eficiente (variância) que o AASc.

Ademais, efetuaremos o cálculo do tamanho amostral de dois modos: o primeiro será considerando as colunas como estratos e o segundo, as linhas. Isso foi feito para que pudéssemos dar opções ao pesquisador, de modo que ele possa nos informar quais desses elementos são de fato os estratos de sua pesquisa.

## Considerando as colunas como estratos

### Abordagem 1: Amostragem Aleatória Simples com reposição (AASc) em conjunto com Amostragem Estratificada Proporcional (AEpr)

Nessa abordagem, inicialmente calcularemos o tamanho total amostral por meio do plano AASc, dado pela fórmula

$$n_{total} = \frac{\sigma_{total}^2}{D}$$

onde  $D = \frac{B^2}{z_\alpha^2}$  é uma função do erro  $B$  escolhido, fixado um nível de confiança  $z_\alpha$  e a variância total é a soma das variâncias de cada estrato, isto é

$$\sigma_{total}^2 = \frac{\sum_{h=1}^H \sigma_h^2}{N}$$

Além disso, a variância de cada estrato é dada pelo plano AE como se segue

$$\sigma_h^2 = \frac{\sum_{i=1}^{N_h}}{N_h}$$

Sobre a  $D$  é calculada a partir de um erro escolhido pelo pesquisador, que chamamos de  $B$  ponderado por um valor dado pelo nível de significância desejado  $\alpha$  que, em geral, é usado o valor 5%. É importante salientar que quanto maior o erro que se está disposto a cometer, maior o tamanho amostral necessário. Contudo, se for possível escolher um erro maior, o tamanho da amostra também diminui consideravelmente. Contudo, ao fazer essa escolha, perde-se precisão em suas análises e inferências. É uma troca que se faz para que seja possível reduzir o tamanho amostral. Tal escolha fica a critério do pesquisador, e sua disposição em trabalhar com uma amostra maior ou menor.

Tendo calculado o tamanho total da amostra pelo plano AASc, conseguimos determinar o tamanho de amostra necessário dentro de cada grupo usando o plano AEpr, dado pela fórmula

$$n_h = n \frac{N_h}{N} = nW_h$$

,

Table 1: Erro e tamanho da amostra

B	tamanhos1
3.770377	33922
3.780378	33743
3.790379	33565
3.800380	33389
3.810381	33214
3.820382	33040

Table 2: Resultados pela abordagem 1

Estratos	Tamanho.de.amostra
Deezer	293
Google Podcasts	26
Soundcloud	41
Spotify	8441
Youtube	23918
Apple Podcasts	91
Listen Notes	211
Bit Chute	320
MGTOW TV	561
Castbox	23

em que  $W_h$  é o peso que cada grupo possui na população.

Por meio de simulação computacional, obtivemos que o erro mínimo para calcular o tamanho da amostra é 3.770377 resultando em um tamanho amostral de 33922. A Table 2 mostra o tamanho amostral dentro de cada grupo, ao passo que a Table 1 mostra a redução do tamanho amostral necessário, a medida que o erro escolhido aumenta.

Por exemplo, para um erro de 10.001000, o tamanho de amostra seria 4822, gerando a tabela abaixo

Table 3: Erro e tamanho da amostra

B	tamanhos2
0.030003	34058
0.040004	34057
0.050005	34054
0.060006	34052
0.070007	34049
0.080008	34045

Estratos	Tamanho.de.amostra
Deezer	42
Google Podcasts	4
Soundcloud	6
Spotify	1200
Youtube	3400
Apple Podcasts	13
Listen Notes	30
Bit Chute	46
MGTOW TV	80
Castbox	4

## Abordagem 2: Amostragem Aleatória Simples sem reposição (AASs) em conjunto com Amostragem Estratificada Proporcional (AEpr)

nessa segunda abordagem, todo o procedimento para calcular o tamanho da amostra é análogo. A única diferença será no cálculo do  $n_{total}$  pois agora ele será determinado pelo plano AASs, dado pela fórmula

$$n_{total} = \frac{1}{\frac{D}{\sigma_{total}^2} + \frac{1}{N}}$$

A Table 4 mostra o tamanho amostral de cada estrato quando usamos o tamanho amostral obtido pelo erro mínimo na Table 3 que é 34058. Esse tamanho pode diminuir se o pesquisador estiver disposto a aumentar o erro.

Tentando diminuir o tamanho da amostra, para um erro de 10.001000, tem-se um tamanho de 4224, gerando a tabela abaixo

Table 4: Resultados pela abordagem 2

Estratos	Tamanho.de.amostra
Deezer	294
Google Podcasts	26
Soundcloud	41
Spotify	8475
Youtube	24014
Apple Podcasts	91
Listen Notes	211
Bit Chute	321
MGTOW TV	563
Castbox	23

Estratos	Tamanho.de.amostra
Deezer	37
Google Podcasts	4
Soundcloud	6
Spotify	1052
Youtube	2979
Apple Podcasts	12
Listen Notes	27
Bit Chute	40
MGTOW TV	70
Castbox	3

### Considerando as linhas como estratos

#### Abordagem 1: Amostragem Aleatória Simples com reposição (AASc) em conjunto com Amostragem Estratificada Proporcional (AEpr)

B	tamanhos1_linha
15.36154	34048
15.37154	34004
15.38154	33960
15.39154	33916
15.40154	33872
15.41154	33828

O erro mínimo encontrado foi 15.36154 resultando em um tamanho de amostra de 34048. Vale a observação de que o uso desse plano amostral, neste caso não é bom, pois são necessários

erros muito grandes para reduzir o tamanho da amostra.

Estratos	Tamanho.de.amostra
À Deriva	226
Abraham	30
Admirável Cast Novo	5
Alpha Essencia	244
Amantecast	1
Arte Ataca!	28
Atitude Alfa	646
BAILE AMALDIÇOADO DO TIÃO	36
Barzinho MGTOW	16
Baú da Real	185

**Abordagem 2: Amostragem Aleatória Simples com reposição (AASs) em conjunto com Amostragem Estratificada Proporcional (AEpr)**

B	tamanhos2_linha
0.090009	34059
0.100010	34059
0.110011	34059
0.120012	34058
0.130013	34058
0.140014	34058

O erro mínimo obtido foi 0.090009 resultando em um tamanho de amostra de 34059

Estratos	Tamanho.de.amostra
À Deriva	226
Abraham	30
Admirável Cast Novo	5
Alpha Essencia	244
Amantecast	1
Arte Ataca!	28
Atitude Alfa	646
BAILE AMALDIÇOADO DO TIÃO	36
Barzinho MGTOW	16
Baú da Real	185