



Departamento de Estatística - EST

Plano Amostral para entrevista de moradores do Acre

Arthur Gonçalves de Souza - 221022794

Técnicas de Amostragem 2

Prof. Alan Ricardo da Silva

Brasília

2025

Sumário

| | Página |
|--|--------|
| 1 Introdução | 1 |
| 2 Referencial teórico | 2 |
| 2.1 Amostragem | 2 |
| 2.2 Amostragem Aleatória Simples | 2 |
| 2.3 Amostragem sistemática | 2 |
| 2.4 Amostragem Aleatória Estratificada | 2 |
| 2.5 Amostragem Aleatória por Conglomerados | 3 |
| 2.6 Amostragem complexa | 3 |
| 2.7 Pesos amostrais | 4 |
| 2.8 Unidades oversized | 4 |
| 2.9 Unidades undersized | 5 |
| 3 Cálculos do tamanho da Amostra | 6 |
| 4 Primeiro Estágio: Municípios | 9 |
| 5 Segundo Estágio: Setores Censitários | 12 |
| 6 Terceiro estágio: Domicílios | 15 |
| 7 Cálculo das estimativas | 15 |
| 7.1 Média e variância | 15 |
| 7.2 Replicação do erro amostral | 17 |
| 8 Referências | 18 |

1 Introdução

O seguinte relatório possui como objetivo principal a elaboração de um plano amostral para entrevistar pessoas em seus domicílios no Estado do Acre. Nesse contexto, a seleção amostral deve envolver certas particularidades dentro do estado, visto que por exemplo, as opiniões de pessoas que vivem regiões rurais e urbanas podem variar bastante, o que nos leva a garantir que essa heterogeneidade esteja presente na amostra final.

Assim, o plano amostral teve de considerar então uma conglomeração com seleção em três estágios, considerando como primeiro estágio a seleção dos municípios, no segundo estágio a seleção de setores censitários, os quais serão previamente estratificados com base na situação do setor (urbano ou rural), e por último o terceiro estágio, que selecionará os domicílios os quais serão visitados para realizar as entrevistas. Esse método de amostragem, que emprega múltiplos estágios, estratificação e pesos desiguais, é conhecido como amostragem complexa.

Para realizar as seleções, foram utilizados dois bancos de dados, sendo primeiro os dados dos setores censitários de 2022, disponibilizados pelo IBGE, e também os dados dos domicílios, disponibilizados pelo CNEFE. Os dados de setores censitários apresentam 2.215 observações, com 36 variáveis, abordando características geográficas como o município que o setor se localiza, situação dele, e também informações como número de habitantes e domicílios no setor. Os dados de domicílios apresentam 410.524 observações e 34 variáveis, englobando informações específicas sobre o endereço dos domicílios. Será utilizado ainda os dados das malhas geográficas dos setores censitários, os quais também são disponibilizados pelo IBGE, para a produção de mapas.

Para o tratamento das bases de dados, elaboração de gráficos, cálculos e seleções de amostra, foi utilizado o software R, versão 4.4.2, com a seed 1 para replicação das análises.

2 Referencial teórico

2.1 Amostragem

A amostragem é uma técnica estatística que permite conseguir resultados aproximados para a população a partir de uma quantidade menor de informações, ou seja, por meio de observações de apenas um “pedaço” dessa população. Dessa forma, consegue-se, com um intervalo de confiança, reduzir os custos e otimizar o tempo de coleta de informações sem perder a credibilidade para o estudo em questão.

2.2 Amostragem Aleatória Simples

Na amostragem aleatória simples, cada componente da população tem a mesma probabilidade de ser selecionado para fazer parte da amostra, ou seja, dada uma população com N indivíduos, cada um possui probabilidade igual a $\frac{1}{N}$ de ser selecionado. Além disso, é necessário que a seleção de indivíduos seja feita de forma aleatória.

2.3 Amostragem sistemática

Na amostragem sistemática, a informação a ser coletada é de fácil acesso (exemplos: assinantes de uma revista, cadastro de funcionários, entre outras situações). De início, determina-se um intervalo de seleção (que pode ser calculado pela divisão da população sobre o tamanho da amostra que será selecionada), o qual é denotado pela letra k .

O primeiro termo da amostra é um elemento entre os k primeiros elementos, sorteado de forma aleatória. O segundo componente da amostra será o k -ésimo elemento seguinte, ou seja, “soma-se” k posições em relação ao primeiro termo amostrado. O terceiro termo será o k -ésimo elemento seguinte, e assim por diante (seguindo uma progressão aritmética de razão k) até que se atinja o número de elementos que se deseja amostrar.

2.4 Amostragem Aleatória Estratificada

Na amostragem aleatória estratificada, a população é dividida em L estratos, em que cada elemento deve pertencer a exatamente um estrato, cada estrato contendo $N_1, N_2, \dots, N_h, \dots, N_L$. Para obter o benefício total da estratificação, o número de elementos em cada estrato deve ser conhecido. Esse método é utilizado para compor grupos mais homogêneos, uma vez que a determinada população é dividida por características distintas. Há três maneiras de realizar a alocação da amostragem aleatória estratificada:

- **Mesmo tamanho (Uniforme):** são retiradas amostras de mesmo tamanho de cada estrato (utiliza-se quando os estratos possuem aproximadamente o mesmo tamanho)
- **Proporcional:** como o próprio nome já diz, realiza-se uma proporção para determinar quantos elementos de cada estrato serão sorteados: quanto maior o estrato, mais elementos serão amostrados daquele estrato
- **Ótima de Neyman:** é o melhor formato, porém aquele que exige mais informações. Considera tanto o tamanho do estrato como também a variância dentro desse estrato

Portanto, se L é o número de estratos, N é o tamanho da população, n é o tamanho total da amostra e N_h é o tamanho do i -ésimo estrato na população, com $h = 1, 2, \dots, L$, então o tamanho de cada estrato será dado por:

- **Mesmo tamanho (Uniforme):** $n_h = \frac{n}{L}$, $h = 1, 2, \dots, L$
- **Proporcional:** $n_h = n \frac{N_h}{N}$, $h = 1, 2, \dots, L$
- **Ótima de Neyman:** para um dado custo fixo, a variância do estimador da média pela amostragem estratificada é minimizada para:

$$n_h = n \frac{W_h S_h}{\sum_{h=1}^L (W_h S_h)}$$

Em que:

- $h = 1, 2, \dots, L$
- W_h : peso do estrato h (N_h/N)
- S_h : variância do estrato h

2.5 Amostragem Aleatória por Conglomerados

Na amostragem aleatória por conglomerados, a população é separada em grupos (*clusters* ou conglomerados). De maneira geral, pode parecer muito semelhante à amostragem estratificada, uma vez que as duas são separadas em blocos, porém as técnicas são contrárias.

Para a amostragem estratificada, os grupos devem ser heterogêneos entre eles, porém homogêneos dentro de cada estrato, e além disso, garante-se na coleta que exista amostra de todos eles. Já na amostragem por conglomerados, os *clusters* são homogêneos entre eles e heterogêneos dentro de cada conglomerado. Assim, espera-se que cada *cluster* tenha as mesmas características da população.

2.6 Amostragem complexa

O termo amostragem complexa é usada quando a amostragem é feita usando conglomerados, estratificação e probabilidades de seleção desiguais (Chambers e Skinner, 2003). Esses procedimentos são empregados para garantir maior eficiência na representatividade da amostra, especialmente em populações grandes e heterogêneas, onde a seleção direta de indivíduos pode ser inviável ou custosa. Esse tipo de abordagem é amplamente utilizado em pesquisas populacionais, levantamentos governamentais e estudos epidemiológicos, nos quais a necessidade de garantir representatividade e minimizar vieses é fundamental.

2.7 Pesos amostrais

Os pesos amostrais são valores atribuídos a cada unidade amostral (indivíduo, domicílio, setor, etc.) para refletir sua representatividade na população total. Eles são essenciais em planos amostrais complexos, onde a seleção das unidades não é equiprovável (ou seja, nem todos os elementos da população têm a mesma chance de serem selecionados). Os pesos são calculados como o inverso da probabilidade de seleção de cada unidade, ajustando as estimativas para que reflitam adequadamente a população de interesse.

A probabilidade de seleção de uma unidade amostral é determinada pelo desenho amostral utilizado (estratificação, conglomerados, seleção proporcional ao tamanho, etc.). Em um plano amostral com múltiplos estágios, o peso final é o produto dos pesos em cada estágio. Por exemplo, em uma amostragem em três estágios (municípios, setores censitários e domicílios), o peso final é dado por:

$$\text{Peso Final} = \frac{1}{P1_i} \times \frac{1}{P2_{ij}} \times \frac{1}{P3_{ijk}}$$

Onde:

- $P1_i$: Probabilidade de seleção do município i no primeiro estágio.
- $P2_{ij}$: Probabilidade de seleção do setor j no município i no segundo estágio.
- $P3_{ijk}$: Probabilidade de seleção do domicílio k no setor j do município i no terceiro estágio.

2.8 Unidades oversized

Um dos principais problemas que podem ser encontrados tratando de seleção de conglomerados de tamanho desigual é a questão das unidades oversizeds ou superdimensionadas. Unidades são definidas dessa forma quando elas apresentam uma grande magnitude com relação à variável de interesse, ocasionando numa probabilidade de seleção extremamente alta, muitas vezes até maior que um. Para identificar essas unidades, antes de realizar o sorteio da amostra, utiliza-se o critério de que um município é considerado oversized caso:

$$Prob > 1 \Leftrightarrow \frac{aT_i}{T_0} > 1 \Leftrightarrow \boxed{T_i > \frac{T_0}{a}}$$

Onde:

- a : Número de unidades a serem selecionadas na amostra;
- T_i : Valor da variável de interesse na unidade em questão;
- T_0 : Total da variável de interesse considerando todas as unidades.

A partir da verificação desse critério, é realizada uma estratificação das unidades, agrupando as unidades oversizeds no estrato 1 (ou estrato “certo”), que portanto, já estarão presentes na amostra devido à sua dimensão, sendo atribuído à essas unidades o peso amostral 1, enquanto as unidades que não foram classificadas como oversizeds, são agrupadas no estrato 2.

Vale destacar também que esse processo é repetido até que não sejam encontradas mais unidades oversizeds. Dado que o processo já foi realizado uma vez, a segunda verificação utilizaria o seguinte critério:

$$T_i > \frac{T_{0II}}{a_{II}}$$

onde:

- a_{II} : Número de unidades a serem selecionadas do estrato 2 (a - número da unidades no estrato 1);
- T_{0II} : Total da variável de interesse considerando todas as unidades do estrato II.

E as repetições vão seguindo nesse sentido, apenas atualizando os valores de a e M_0 .

2.9 Unidades undersized

Outro problema que também pode ser encontrado em seleções de conglomerados de tamanho desigual é basicamente o contrário das unidades oversizeds, agora as unidades undersizeds ou subdimensionadas. Unidades são definidas undersizeds quando estas possuem uma quantidade muito pequena de unidades secundárias, que comporão o segundo estágio. Para realizar tal verificação, após o sorteio das a unidades, analisa-se se alguma unidade se enquadra no seguinte critério:

$$M_i < m_i$$

onde M_i é o número total de unidades secundárias na unidade primária, enquanto que m_i é o número de unidades secundárias que deverão ser selecionadas a partir da unidade primária, e possui a seguinte fórmula:

$$m_i = \frac{a \times \bar{m} \times M_i}{\sum_{i=1}^a M_i}$$

A partir da verificação desse critério, caso haja alguma unidade undersized, é possível juntar esse unidade com alguma outra que seja considerada mais próxima, ou então, caso haja mais de uma, podendo agrupar todas as undersizeds como uma só, para então criar unidades com tamanhos suficientes. Ao contrário da verificação de oversizeds, esse processo só precisa ser realizado uma vez por estágio.

3 Cálculos do tamanho da Amostra

O plano amostral em questão, definido como complexo, será composto por 3 estágios, sendo respectivamente a seleção de municípios, setores censitários e domicílios, considerando pesos desiguais em cada estágio. Para o cálculo dos tamanhos das amostras de cada estágio, será utilizada a estrutura com base no P (rate of homogeneity - roh), onde $P = \frac{Def f - 1}{\bar{P} - 1}$. Nesse contexto, \bar{P} representa o número médio de domicílios por setor censitários, e com base nos dados de setores, $\bar{P} \approx 150,11$. Assim, de acordo com o demandante, foi fixado um valor de $Def f = 2,5$, e então obtém-se o valor de P :

$$P = \frac{Def f - 1}{\bar{P} - 1} = \frac{2,5 - 1}{150,11 - 1} \approx 0,01$$

A partir desse valor de P , pode-se agora calcular o valor de \bar{p} , que representa o número médio de domicílios que devem ser sorteados de cada setor censitário na amostra. Para tal, será necessário utilizar os custos reservados para o projeto, que de acordo com o demandante, são:

- $C = R\$140.000$, sendo o custo total da pesquisa;
- $C_0 = R\$24.000$, sendo o custo fixo;
- $C_1 = R\$600,00$, sendo o custo para selecionar um município;
- $C_2 = R\$12,00$, sendo o custo para selecionar um setor censitário;
- $C_3 = R\$10,00$, sendo o custo para entrevistar um domicílio.

Dessa forma, tem-se que:

$$\bar{p} = \sqrt{\frac{(1 - P) C_2}{P C_3}} = \sqrt{\frac{(1 - 0,01) 12}{0,01 10}} \approx 11$$

Seguindo assim, pode-se calcular o valor de \bar{m} , que representa o número médio de setores que devem ser selecionados de cada município na amostra. A fórmula para obtenção de \bar{m} é:

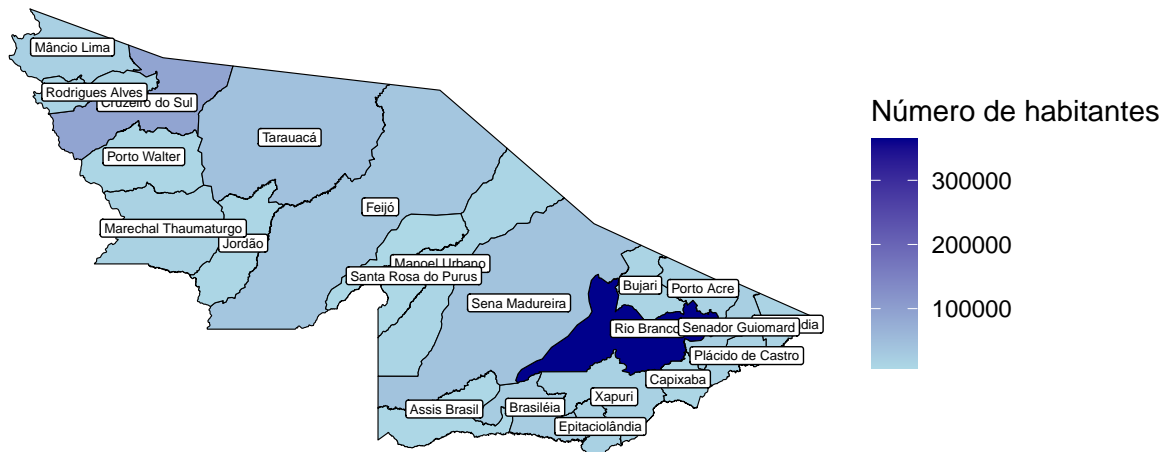
$$\bar{m} = \frac{z_{\frac{\alpha}{2}}^2 \times CV^2 \times (P + \frac{1-P}{\bar{p}})}{r^2}$$

Com:

- $z_{\frac{\alpha}{2}}$: quantil da distribuição normal padrão e aproximadamente igual a 1,96 para $\alpha = 5\%$;
- α : nível de significância;
- CV : Coeficiente de variação da variável analisada
- r : Erro percentual
- P : rate of homogeneity
- \bar{p} : Número médio de domicílios que devem ser sorteados de cada setor censitário na amostra

Porém, não há um CV já definido para o estudo, e nesse caso o estimaremos com base na proporção de municípios grandes (número de habitantes maior que 50 mil) no Acre, que chamaremos de prop:

Número de habitantes por municípios do Acre



Fonte: Elaboração própria

Tabela 1: Número de habitantes por municípios do Acre

| Município | Número de habitantes | Classificação |
|----------------------|----------------------|---------------|
| Acrelândia | 14.021 | Pequeno |
| Assis Brasil | 8.100 | Pequeno |
| Brasília | 26.000 | Pequeno |
| Bujari | 12.917 | Pequeno |
| Capixaba | 10.392 | Pequeno |
| Cruzeiro do Sul | 91.888 | Grande |
| Epitaciolândia | 18.757 | Pequeno |
| Feijó | 35.426 | Pequeno |
| Jordão | 9.222 | Pequeno |
| Manoel Urbano | 11.996 | Pequeno |
| Marechal Thaumaturgo | 17.093 | Pequeno |
| Mâncio Lima | 19.294 | Pequeno |
| Plácido de Castro | 16.560 | Pequeno |
| Porto Acre | 16.693 | Pequeno |
| Porto Walter | 10.735 | Pequeno |
| Rio Branco | 364.756 | Grande |
| Rodrigues Alves | 14.938 | Pequeno |
| Santa Rosa do Purus | 6.723 | Pequeno |
| Sena Madureira | 41.343 | Pequeno |
| Senador Guiomard | 21.454 | Pequeno |
| Tarauacá | 43.467 | Pequeno |
| Xapuri | 18.243 | Pequeno |

Tabela 2: Proporção de municípios grandes no Acre

0,0909091

A partir desse valor, obtém-se $CV = \sqrt{prop * (1 - prop)} \approx 0,287$, e então podemos gerar uma tabela com os valores possíveis de \bar{m} com base em valores r fixados. Consequentemente, também é possível obter os possíveis valores a , que representa o número de municípios a serem selecionados na amostra, e é calculado como:

$$a = \frac{C - C_0}{C_1 + C_2 \times \bar{m} + C_3 \times \bar{m} \times \bar{p}}$$

que é derivada da função de custos $C = C_0 + C_1 \times a + C_2 \times a \times \bar{m} + C_3 \times a \times \bar{m} \times \bar{p}$

A partir dessas fórmulas, é obtida a seguinte tabela:

Tabela 3: Tabela com a relação de erros fixados e os tamanhos de amostra

| Erros | Setores | Municípios |
|-------|---------|------------|
| 1% | 318 | 2 |
| 1,5% | 141 | 6 |
| 2% | 79 | 10 |
| 2,5% | 51 | 14 |
| 3% | 35 | 20 |
| 3,5% | 26 | 26 |
| 4% | 20 | 32 |
| 4,5% | 16 | 38 |
| 5% | 13 | 45 |

A partir da tabela acima, e em concordância com o demandante, optou-se por definir $\bar{m} = 51$ e $a = 14$, concluindo portanto, que a amostra final apresentará:

Tabela 4: Quantidade de Unidades no Plano Amostral

| Unidades | Quantidade |
|------------|------------|
| Municípios | 14 |
| Setores | 714 |
| Domicílios | 7.854 |

4 Primeiro Estágio: Municípios

Como objetivo do plano amostral é representar toda a população do estado do Acre, é importante se basear nesse número para verificar posteriormente se a amostra realmente está sendo representativa, a partir dos pesos amostrais. Nesse sentido, segue abaixo a população total do Estado do Acre:

Tabela 5: População Acre (2022)

830.018 pessoas

O primeiro passo para a seleção da amostra dos municípios, nesse caso o 1º estágio, é a verificação de quais municípios são considerados oversized. O processo teve de ser realizado 3 vezes, e ao final foi obtido o seguinte resultado:

Tabela 6: Análise de municípios oversized

| Município | Número de habitantes | Estrato | Oversized |
|----------------------|----------------------|---------|-----------|
| Rio Branco | 364.756 | 1 | Sim |
| Cruzeiro do Sul | 91.888 | 1 | Sim |
| Tarauacá | 43.467 | 1 | Sim |
| Sena Madureira | 41.343 | 1 | Sim |
| Feijó | 35.426 | 1 | Sim |
| Brasiléia | 26.000 | 2 | Não |
| Senador Guiomard | 21.454 | 2 | Não |
| Mâncio Lima | 19.294 | 2 | Não |
| Epitaciolândia | 18.757 | 2 | Não |
| Xapuri | 18.243 | 2 | Não |
| Marechal Thaumaturgo | 17.093 | 2 | Não |
| Porto Acre | 16.693 | 2 | Não |
| Plácido de Castro | 16.560 | 2 | Não |
| Rodrigues Alves | 14.938 | 2 | Não |
| Acrelândia | 14.021 | 2 | Não |
| Bujari | 12.917 | 2 | Não |
| Manoel Urbano | 11.996 | 2 | Não |
| Porto Walter | 10.735 | 2 | Não |
| Capixaba | 10.392 | 2 | Não |
| Jordão | 9.222 | 2 | Não |
| Assis Brasil | 8.100 | 2 | Não |
| Santa Rosa do Purus | 6.723 | 2 | Não |

A partir da análise de oversized, segue-se então para o sorteio dos municípios do estrato 2. Restaram-se 17 municípios no estrato 2, dos quais devem-se selecionar 9 ($a = 14$ - municípios oversized = 5). Para realizar a seleção foi optado por utilizar o método sistemático.

Tabela 7: Amostra de municípios

| | Municípios | Número de habitantes | Estrato |
|----|-------------------|----------------------|---------|
| 1 | Rio Branco | 364.756 | 1 |
| 2 | Cruzeiro do Sul | 91.888 | 1 |
| 3 | Tarauacá | 43.467 | 1 |
| 4 | Sena Madureira | 41.343 | 1 |
| 5 | Feijó | 35.426 | 1 |
| 6 | Porto Acre | 16.693 | 2 |
| 7 | Brasiléia | 26.000 | 2 |
| 8 | Senador Guimard | 21.454 | 2 |
| 9 | Bujari | 12.917 | 2 |
| 10 | Mâncio Lima | 19.294 | 2 |
| 11 | Xapuri | 18.243 | 2 |
| 12 | Acrelândia | 14.021 | 2 |
| 13 | Rodrigues Alves | 14.938 | 2 |
| 14 | Plácido de Castro | 16.560 | 2 |

Em seguida, com a amostra de municípios já definida, pode-se realizar então a análise de municípios undersizeds, realizando o cálculo do valor de m_i para cada município selecionado, e verificando se algum possui o número de setores menor que m_i

Tabela 8: Análise de municípios undersizeds

| Municípios | Número de setores | m_i | Undersized |
|-------------------|-------------------|-------|------------|
| Rio Branco | 667 | 282 | Não |
| Cruzeiro do Sul | 188 | 79 | Não |
| Tarauacá | 132 | 56 | Não |
| Sena Madureira | 135 | 57 | Não |
| Feijó | 154 | 65 | Não |
| Porto Acre | 42 | 18 | Não |
| Brasiléia | 66 | 28 | Não |
| Senador Guimard | 48 | 20 | Não |
| Bujari | 33 | 14 | Não |
| Mâncio Lima | 61 | 26 | Não |
| Xapuri | 53 | 22 | Não |
| Acrelândia | 29 | 12 | Não |
| Rodrigues Alves | 41 | 17 | Não |
| Plácido de Castro | 42 | 18 | Não |

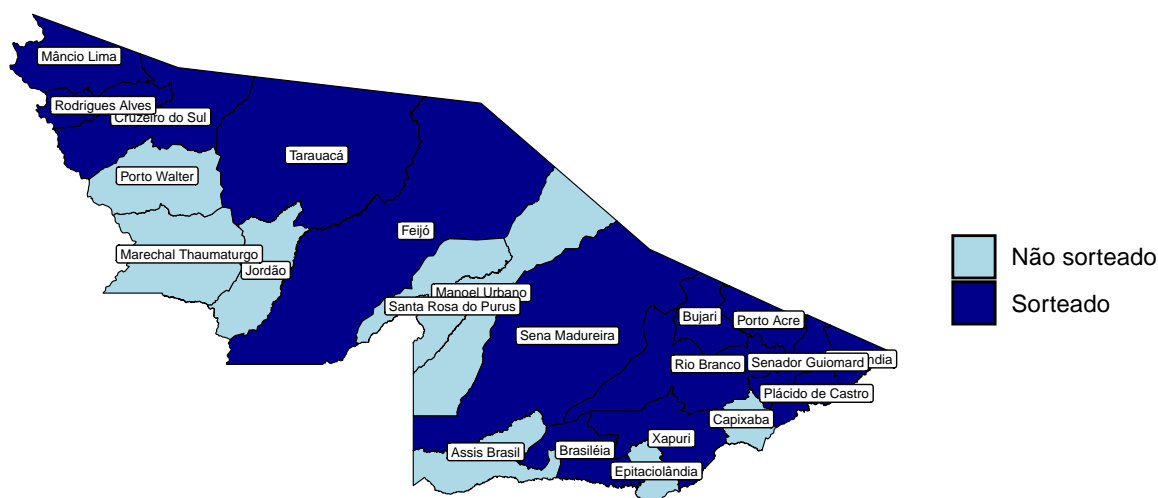
Observando a tabela acima, percebe-se que não foram identificados municípios undersizeds na amostra. Vale ressaltar que como o plano amostral é proporcional ao tamanho de cada unidade, torna-se bem difícil obter unidades undersizeds, visto que o cálculo do m_i evita que isso aconteça. Sendo assim, para finalizar a seleção, calcula-se os pesos amostrais para cada município, cuja soma do número de habitantes ponderada deve coincidir com a população total do estado do Acre.

Tabela 9: Soma do total de habitantes ponderados pelo peso no 1º estágio

830.018

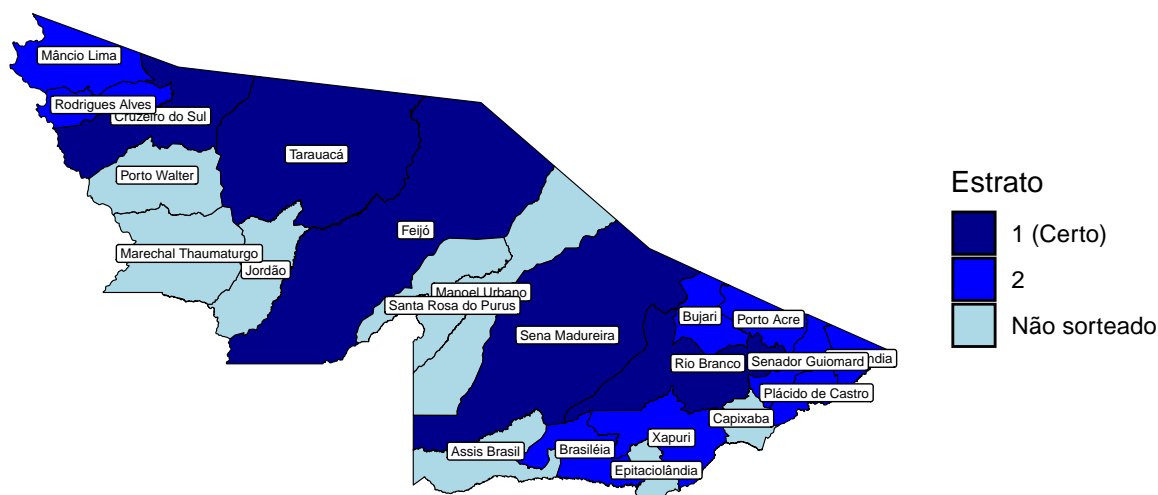
Como o resultado coincide com a população do estado, pode-se afirmar que a seleção no 1º estágio foi realizada de forma adequada. Sendo assim, seguem abaixo dois mapas representando os municípios selecionados para a amostra.

Mapa da amostra de municípios do Acre



Fonte: Elaboração própria

Mapa da amostra de municípios do Acre (por estrato)



Fonte: Elaboração própria

5 Segundo Estágio: Setores Censitários

Ao tratar do segundo estágio, que objetiva a seleção de setores censitários, percebe-se um possível fator de heterogeneidade, relacionado com a situação do setor, categorizado em rural e urbano. Nesse sentido, é ideal que o segundo estágio seja estratificado com base em tal variável. Para tal, optou-se pela alocação proporcional, fazendo com que se tenha que dividir o m_i de cada município com relação à proporção de cada categoria no mesmo. Seguem abaixo o gráfico da proporção dos setores rurais e urbanos por município, além da tabela com as devidas alocações.

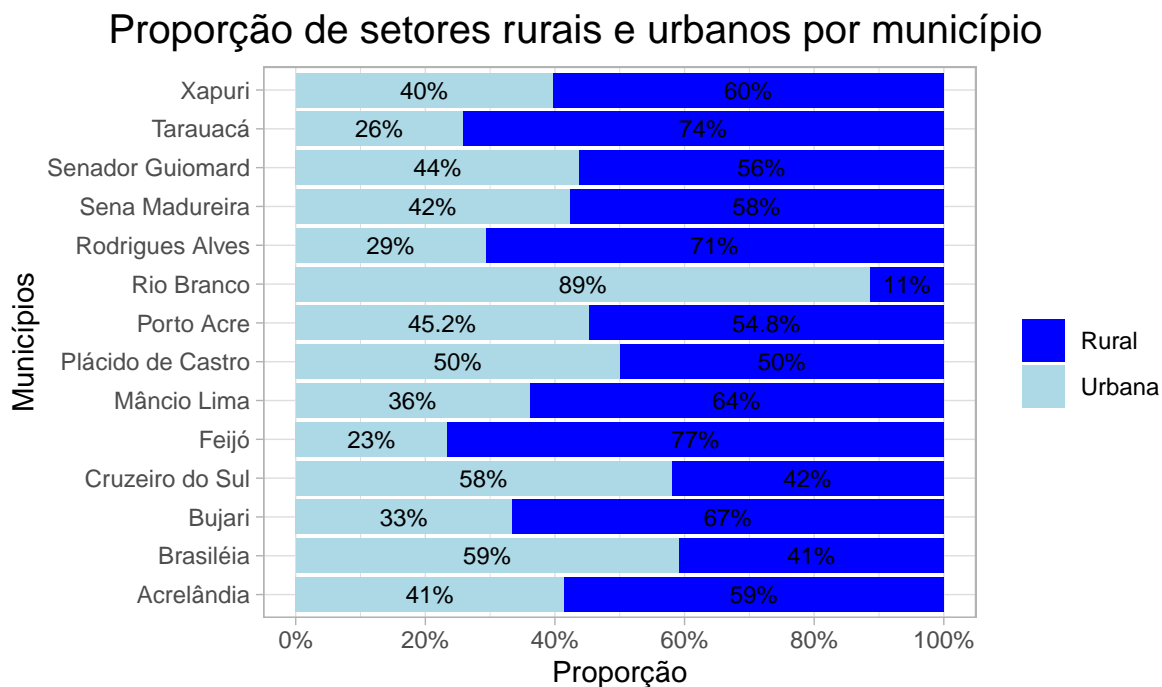


Tabela 10: Setores por municípios selecionados

| Municípios | Número de setores | Setores rurais | Setores urbanos | m_i | $m_i Rural$ | $m_i Urbano$ |
|-------------------|-------------------|----------------|-----------------|-------|-------------|--------------|
| Acrelândia | 29 | 17 | 12 | 12 | 7 | 5 |
| Brasiléia | 66 | 27 | 39 | 28 | 11 | 17 |
| Bujari | 33 | 22 | 11 | 14 | 9 | 5 |
| Cruzeiro do Sul | 188 | 79 | 109 | 79 | 33 | 46 |
| Feijó | 154 | 118 | 36 | 65 | 50 | 15 |
| Mâncio Lima | 61 | 39 | 22 | 26 | 17 | 9 |
| Plácido de Castro | 42 | 21 | 21 | 18 | 9 | 9 |
| Porto Acre | 42 | 23 | 19 | 18 | 10 | 8 |
| Rio Branco | 667 | 76 | 591 | 282 | 32 | 250 |
| Rodrigues Alves | 41 | 29 | 12 | 17 | 12 | 5 |
| Sena Madureira | 135 | 78 | 57 | 57 | 33 | 24 |
| Senador Guimard | 48 | 27 | 21 | 20 | 11 | 9 |

| Municípios | Número de setores | Setores rurais | Setores urbanos | m_i | m_iRural | $m_iUrbano$ |
|------------|-------------------|----------------|-----------------|-------|------------|-------------|
| Tarauacá | 132 | 98 | 34 | 56 | 42 | 14 |
| Xapuri | 53 | 32 | 21 | 22 | 13 | 9 |

Após a alocação proporcional, parte-se então para a análise dos setores censitários oversizeds. Agora no segundo estágio, a dinâmica muda um pouco, e serão considerados:

- T_i : Número de habitantes no setor censitário.
- T_0 : Número total de habitantes dos setores naquele estrato (rural ou urbano) para aquele município;
- a : m_{ih} , ou seja, número de setores censitários a serem selecionados do h-ésimo estrato (rural ou urbano) do i-ésimo município.

A partir da análise de setores oversizeds, foi percebido um possível empecilho para a amostragem, 7 setores que foi considerados como oversizeds, ou seja, estavam no estrato certo, não continham domicílios cadastrados nos dados do CNEFE, o que levou a desconsiderar esses setores na amostra, e então forçar o sorteio a não selecioná-los. Dessa forma, sumariza-se da seguinte forma a análise de setores oversizeds.

Tabela 11: Análise de setores oversizeds

| Municípios | Setores rurais | Setores urbanos | m_i Rural | m_i Urbano | Setores rurais oversizeds | Setores urbanos oversizeds |
|-------------------|----------------|-----------------|----------------|-----------------|---------------------------|----------------------------|
| Acrelândia | 17 | 12 | 7 | 5 | 0 | 0 |
| Brasiléia | 27 | 39 | 11 | 17 | 0 | 0 |
| Bujari | 22 | 11 | 9 | 5 | 0 | 1 |
| Cruzeiro do Sul | 79 | 109 | 33 | 46 | 5 | 0 |
| Feijó | 118 | 36 | 50 | 15 | 28 | 0 |
| Mâncio Lima | 39 | 22 | 17 | 9 | 7 | 0 |
| Plácido de Castro | 21 | 21 | 9 | 9 | 0 | 0 |
| Porto Acre | 23 | 19 | 10 | 8 | 0 | 0 |
| Rio Branco | 76 | 591 | 32 | 250 | 0 | 0 |
| Rodrigues Alves | 29 | 12 | 12 | 5 | 0 | 0 |
| Sena Madureira | 78 | 57 | 33 | 24 | 12 | 0 |
| Senador Guiomard | 27 | 21 | 11 | 9 | 0 | 0 |
| Tarauacá | 98 | 34 | 42 | 14 | 16 | 0 |
| Xapuri | 32 | 21 | 13 | 9 | 1 | 1 |

A partir de tal análise, segue-se então para o sorteio dos setores censitários não oversizeds dentro de situação para cada município, lembrando que serão sorteados m_{ih} setores de cada estrato e de município. Nesse estágio, a seleção foi feita por meio de seleção aleatória.

Após a seleção dos 714 setores, pode-se então verificar quais setores podem ser definidos como undersizeds. Nesse estágio, calcula-se o valor de p_{ihj} , que representa o número de domicílios que serão selecionados de cada setor selecionado na amostra. Nesse contexto, calcula-se p_{ijh} da seguinte forma:

$$p_{ihj} = \frac{m_{ih} \times \bar{p} \times P_{ihj}}{\sum_{j=1}^{m_{ih}} P_{ihj}}$$

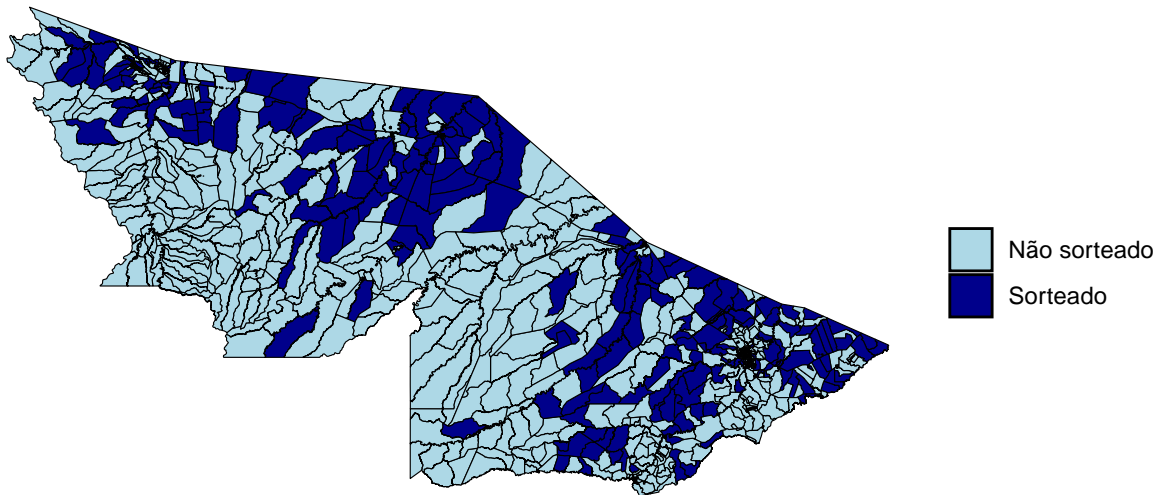
Através da análise de setores undersizeds, não encontrou-se nenhum setor nessa situação, novamente devido ao formato de cálculo de p_{ihj} , que dificulta que tal situação aconteça. Sendo Assim, parte-se então para o cálculo dos pesos amostrais para cada setor, cuja soma do número de habitantes ponderada por este peso também deve coincidir com a população total do estado do Acre.

Tabela 12: Soma do total de habitantes ponderados pelo peso no 2º estágio

| |
|---------|
| 830.018 |
|---------|

Dessa forma, como o resultado coincide com a população total do estado do Acre, pode-se afirmar que a seleção no 2º estágio foi realizada de forma adequada. Sendo assim, Segue abaixo o mapa representando os setores que foram selecionados para a amostra.

Mapa da amostra de setores censitários do Acre



Fonte: Elaboração própria

6 Terceiro estágio: Domicílios

A partir da seleção dos setores que comporão a amostra, pode-se então partir para o 3º estágio do plano amostral, que selecionará agora p_{ihj} domicílios de cada setor selecionado. Porém, após o cálculo dos p_{ihj} , percebeu-se que alguns apresentaram valor menor que 0, e portanto, foram arredondados para o valor de 1, para que ao menos seja selecionado um município de cada setor. Neste sentido, o número de domicílios na amostra aumentou de 7.854 para 7.879, o que não é um problema, visto que o custo real do projeto está em R\$137.508,00, mais de 2 mil reais a menos que o custo reservado para o projeto, o que permite a alocação de verba para a coleta dos 25 domicílios “extras”, alcançando então o custo real de R\$137.758,00.

Sendo assim, utilizando a base de dados do CNEFE, seleciona-se por meio de sorteio aleatório p_{ihj} domicílios de cada setor selecionado no 2º estágio, totalizando então 7.879 municípios.

Após a seleção, é de interesse calcular o peso final de cada domicílio selecionado, obtido pela multiplicação dos pesos de cada estágio. A soma do número de pessoas por domicílio ponderada pelo peso final, deve resultar então no total de habitantes no estado do Acre, que é exatamente o que acontece:

Tabela 13: Soma do total de habitantes por domicílio ponderado pelo peso final

| |
|---------|
| 830.018 |
|---------|

7 Cálculo das estimativas

7.1 Média e variância

Quando utiliza-se um desenho amostral complexo, com probabilidades de seleção desiguais, a maioria das estatísticas de interesse não são simples funções lineares dos dados observados, e por isso são requeridos métodos alternativos para obtenção das estimativas, pois caso contrário, até as estimativas pontuais serão viesadas. Dentre as abordagens possíveis, existem 2 mais comuns, que são:

- Através de linearização por série de Taylor (TSL);
- Através do uso de procedimentos de reamostragem para estimação da variância, como JRR (Jackknife Repeated Replication), BRR (Balanced Repeated Replication) e Bootstrap.

No contexto em questão, para obter a estimativa da variância, optou-se por utilizar o método de Bootstrap. O método consiste na simulação de R réplicas da amostra, selecionando uma AAS de $a_h - 1$ UPA's por estrato, calculando novos pesos para cada réplica, e obtendo então as estimativas ponderadas da média populacional \bar{y}_w , com $r = 1, \dots, R$ réplicas. A partir dessa estimativa, obtém-se a estimativa da variância amostral, calculada como:

$$\widehat{var}_{boot}(\bar{y}_w) = \frac{1}{r-1} \sum_{r=1}^R (\bar{y}_r - \bar{y}_w)^2$$

onde:

$$\bar{y}_w = \frac{\sum_{i=1}^n y_i w_i}{\sum_{i=1}^n w_i}, \quad \bar{y}_r = \frac{\sum_{j=1}^n y_j w_{j,rep}}{\sum_{j=1}^n w_{j,rep}}, \quad w_{j,rep} = w_j \times \frac{ah}{ah-1} m_{hi}(r)$$

e m_{hi} é o número de vezes que a UPA_i do estrato h foi selecionada na réplica r .

Dado a natureza do estudo, para a estimação da variância, foi simulada uma variável aleatória bernoulli com o probabilidade $p = 0,0909091$, que é a proporção de municípios grandes no Acre, estimativa que já foi utilizada anteriormente. Após a simulação, obteve-se a seguinte distribuição da variável

Tabela 14: Distribuição da variável simulada

| Variável simulada | Frequência |
|-------------------|------------|
| 0 | 7.121 |
| 1 | 758 |

Tabela 15: Estimativa da média da variável simulada

0,0965673

Dessa forma, obteve-se a seguinte estimativa para a variância via Bootstrap com 1000 replicações:

Tabela 16: Variância estimada por replicação via Bootstrap

0,000016

Tabela 17: Intervalo de confiança para a média via replicação Bootstrap

| Limite inferior | Estimativa | Limite superior |
|-----------------|------------|-----------------|
| 0,08997 | 0,09657 | 0,10316 |

7.2 Replicação do erro amostral

Além disso, também foi realizada a replicação do erro amostral por meio da variável simulada, obtendo as estimativas da média e do erro padrão para cada município, conforme disponibilizado abaixo:

Tabela 18: Replicação do erro amostral por município

| Municípios | Variável simulada | Erro |
|-------------------|-------------------|--------|
| Acrelândia | 0,0727 | 0,0224 |
| Brasiléia | 0,1120 | 0,0191 |
| Bujari | 0,0473 | 0,0227 |
| Cruzeiro do Sul | 0,1036 | 0,0128 |
| Feijó | 0,0728 | 0,0125 |
| Mâncio Lima | 0,0839 | 0,0229 |
| Plácido de Castro | 0,0802 | 0,0166 |
| Porto Acre | 0,1437 | 0,0207 |
| Rio Branco | 0,0942 | 0,0056 |
| Rodrigues Alves | 0,0965 | 0,0209 |
| Sena Madureira | 0,1169 | 0,0227 |
| Senador Guiomard | 0,0975 | 0,0176 |
| Tarauacá | 0,1031 | 0,0176 |
| Xapuri | 0,1347 | 0,0351 |

Percebe-se portanto que as estimativas dos erros padrão para cada município apresentado estão bem próximas do erro fixado nos cálculos de tamanho da amostra, de 2,5%, havendo apenas a ocorrência do município Xapuri com o erro padrão maior que 0,025.

8 Referências

SILVA, ALAN. **Técnicas de Amostragem 2**. Brasília: Universidade de Brasília 20 jan 2025. Notas de aula.

Downloads | IBGE. Disponível em: https://www.ibge.gov.br/estatisticas/downloads-estatisticas.html?caminho=Censos/Censo_Demografico_2022/Agregados_por_Setores_Censitarios/Agregados_por_Setor_xlsx. Acesso em: 18 dez 2024.

Malhas Territoriais | IBGE. Disponível em: https://geoftp.ibge.gov.br/organizacao_do_territorio/malhas_territoriais/malhas_de_setores_censitarios__divisoes_intramunicipais/censo_2022/setores/shp/UF/. Acesso em: 18 dez. 2024.

Cadastro Nacional de Endereços para Fins Estatísticos | IBGE. Disponível em: <https://www.ibge.gov.br/estatisticas/sociais/populacao/38734-cadastro-nacional-de-enderecos-para-fins-estatisticos.html?edicao=40122&t=downloads>. Acesso em: 18 dez. 2024.

CHAMBERS, R. L.; SKINNER, C. J. Analysis of survey data. Chichester (England) ; Hoboken (Nj): J. Wiley, 2003.