

Plano Amostral

Felipe Barletta

July 12, 2016

1 Amostra por domicílios

O objetivo do estudo, que é verificar o comportamento acerca do descarte de resíduos de cozinha adequadamente no bairro de Santa Felicidade em Curitiba-PR.

Como há uma limitação operacional para a coleta dos dados, optou-se por simular vários cenários de amostragem para que a pesquisadora adote o mais viável planejamento amostral dentro de suas condições de aplicação dos questionários e captação das informações.

O planejamento amostral será realizado com a premissa de uma amostra por domicílio dentro da região geográfica que abrange a população alvo da pesquisa.

A população alvo são todos os domicílios residenciais do bairro de Santa Felicidade em Curitiba-PR. Desta área será coletada a amostra do estudo.

Portanto a unidade amostral será a residência e a unidade observacional será a pessoa entrevistada(responsável pelo domicílio).

Serão simulados dois cenários:

- Amostragem Estratificada
- Amostragem aleatória simples sem reposição

Em cada cenário, o tamanho da amostra será alocado com diferentes erros amostrais:

- 0,03
- 0,05
- 0,06

- 0,08
- 0,10

1.1 Estimativa dos parâmetros

Por convenção denotamos N para o tamanho populacional e n o tamanho amostral.

Antes da descrição dos tipos de amostragem usados neste estudo, vamos definir e estimar os parâmetros de interesse neste estudo, para isso consideraremos os resultados de uma estudo piloto realizado em um colégio estadual situado no bairro de Santa Felicidade, sobre o descarte de óleo residual de cozinha. Na questão 3 usada para coleta dos dados da amostra piloto, perguntava-se:

”3) Como sua família descarta o resíduo de óleo após seu uso?”

Vamos discutir a estimação de uma proporção P , ou seja, a proporção de respostas da pergunta acima que disseram que armazenam o óleo residual para entrega em ponto de coleta. Quem respondeu diferente desta resposta será considerada a estimativa da proporção complementar, Q , ou $1 - P$.

Dos 254 questionários que compuseram a amostra piloto, foram excluídos, para cálculo das estimativas, 60 que responderam ”Não sei” e 1 que não respondeu. Dos entrevistados 68 responderam que separam o óleo residual para coleta, portanto as estimativas ficaram da seguinte forma:

$$\hat{P} = 68/184 = 0,37$$

$$\hat{Q} = 1 - \hat{P} = 0,63$$

Utilizando a aproximação de normalidade assintótica teremos o seguinte intervalo de confiança para a estimativa pontual de \hat{P} :

$$IC(\hat{P}) = \left(\hat{P} - z_{\alpha} \sqrt{(1-f) \frac{\hat{P}\hat{Q}}{n-1}}; \hat{P} + z_{\alpha} \sqrt{(1-f) \frac{\hat{P}\hat{Q}}{n-1}} \right) \quad (1)$$

Como ainda não sabemos o tamanho da nossa população vamos considerar nosso N , que se refere ao tamanho populacional, como infinito. Então a fração amostral $f = n/N$ tenderá a zero. Fazendo os cálculo teremos um intervalo de confiança com 95% de confiança igual à:

$$IC(\hat{P}) = (0.3; 0.44)$$

1.2 Amostragem Estratificada

Este será o primeiro cenário para o plano amostral.

A amostragem estratificada consiste na divisão de uma população em grupos ou estratos segundo alguma característica conhecida na população em estudo.

Neste estudo ela será usada para a melhoria da precisão da estimativa da proporção de famílias que descartam corretamente o resíduo de cozinha.

Os estratos considerados serão os setores censitários definidos pelo IBGE no último censo de 2010, <http://www.censo2010.ibge.gov.br/sinopseporsetores/?nivel=st>.

Dentro de cada estrato será realizada uma amostra aleatória simples sem reposição, em que cada domicílio dentro dos estratos têm a mesma chance de ser escolhidos. Para cálculo do tamanho da amostra. O bairro de Santa Felicidade possui 38 setores censitários, portanto teremos 38 estratos, como vemos na figura 1.

A expressão utilizada para determinar o tamanho na amostra, considerando estratos com população finita, será:

$$n_h = \frac{N\hat{P}\hat{Q}Z^2}{\hat{P}\hat{Q}Z^2 + (N + 1)E^2} \quad (2)$$

Em que:

- n_h : tamanho da amostra do estrato h ;
- Z : Quantil da distribuição (nível de significância) *Normal*;
- E : Erro amostral assumido.

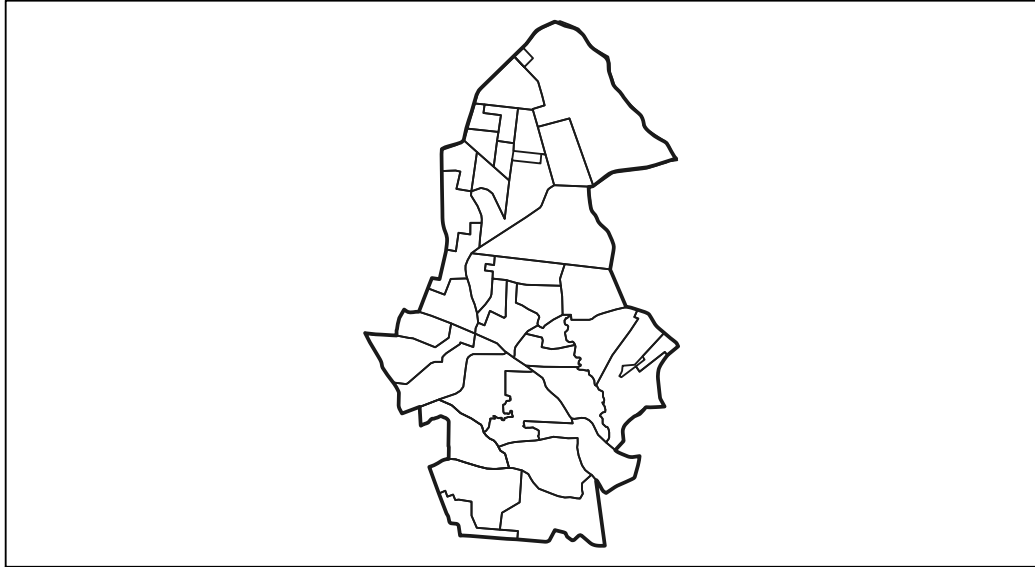


Figure 1: Bairro de Santa Felicidade por setores censitários

1.2.1 Variando erro amostral

Cálculo do tamanho da amostra com erro amostral 0,03, considerando o limite inferior do intervalo de confiança para \hat{P} .

n1	n2	n3	n4	n5	n6	n7	n8	n9	n10	n11	n12	n13	n14	n15	n16	n17	n18
232	227	275	257	217	176	178	229	232	205	261	241	219	303	237	265	204	329
n19	n20	n21	n22	n23	n24	n25	n26	n27	n28	n29	n30	n31	n32	n33	n34	n35	n36
276	188	246	329	197	247	174	159	169	221	235	333	271	359	283	204	214	183
n37	n38																
151	343																
[1]	9069																

Cálculo do tamanho da amostra com erro amostral 0,05, considerando o limite inferior do intervalo de confiança para \hat{P} .

n1	n2	n3	n4	n5	n6	n7	n8	n9	n10	n11	n12	n13	n14	n15	n16	n17	n18
156	154	175	167	150	129	130	155	156	144	169	160	150	186	159	170	143	195
n19	n20	n21	n22	n23	n24	n25	n26	n27	n28	n29	n30	n31	n32	n33	n34	n35	n36
175	135	162	195	140	163	127	119	125	151	158	196	173	205	178	143	148	132
n37	n38																
115	200																

Cálculo do tamanho da amostra com erro amostral 0,06, considerando o limite inferior do intervalo de confiança para \hat{P} .

n1	n2	n3	n4	n5	n6	n7	n8	n9	n10	n11	n12	n13	n14	n15	n16	n17	n18
129	127	141	136	124	109	110	128	129	120	137	132	125	148	130	138	120	154
n19	n20	n21	n22	n23	n24	n25	n26	n27	n28	n29	n30	n31	n32	n33	n34	n35	n36
141	114	133	154	117	133	109	103	107	126	130	155	140	160	143	120	123	112
n37	n38																
99	157																

Cálculo do tamanho da amostra com erro amostral 0,08, considerando o limite inferior do intervalo de confiança para \hat{P} .

n1	n2	n3	n4	n5	n6	n7	n8	n9	n10	n11	n12	n13	n14	n15	n16	n17	n18	n19	n20
89	88	95	93	87	79	80	89	89	85	93	90	87	98	90	94	85	100	95	82
n21	n22	n23	n24	n25	n26	n27	n28	n29	n30	n31	n32	n33	n34	n35	n36	n37	n38		
91	101	83	91	79	76	78	88	90	101	94	103	96	85	86	81	74	102		

Cálculo do tamanho da amostra com erro amostral 0,10, considerando o limite inferior do intervalo de confiança para \hat{P} .

n1	n2	n3	n4	n5	n6	n7	n8	n9	n10	n11	n12	n13	n14	n15	n16	n17	n18	n19	n20
64	63	67	66	63	59	59	64	64	62	66	64	63	68	64	66	62	69	67	60
n21	n22	n23	n24	n25	n26	n27	n28	n29	n30	n31	n32	n33	n34	n35	n36	n37	n38		
65	69	61	65	58	57	58	63	64	70	66	71	67	62	62	60	56	70		

1.3 Amostragem Aleatória Simples Sem Reposição

A amostragem aleatória simples sem reposição (AASs) opera da seguinte forma e não há estratos, ou seja, os setores censitários não são considerados:

- A população será do bairro será numerada de 1 a N ;
- Todos os elementos da população têm a mesma chance de serem sorteados;

- Sorteia-se um elemento de forma aleatória;
- O próximo elemento é sorteado sendo que o anterior foi retirado da população

1.3.1 Variando o erro amostral

Primeiro vamos calcular o tamanho da amostra como no cenário anterior, considerando vários erros amostrais. Sabemos que a população total do bairro é:

$$N = 12174 \text{ domicílios}$$

	n
Erro(0,03)	835
Erro(0,05)	314
Erro(0,06)	220
Erro(0,08)	125
Erro(0,10)	80

1.3.2 Variando o tamanho da amostra

Aqui o tamanho da amostra será fixo e o erro amostral é calculado com base neste tamanho. Os valores fixos da amostra serão, 100, 300, 500 e 1000.

	Erro
n=100	0.089
n=300	0.051
n=500	0.039
n=1000	0.027

2 Cálculo da Variância e Efeito do Planejamento

Cálculo da variância para o cenário da amostragem aleatória simples sem reposição consiedrando erro amostral de 0,03.

Um estimador não viciado (ou não-tendencioso), para $var(\hat{P})$ é dado pela expressão a

seguir: (mais sobre propriedades dos estimadores veja: <http://www.leg.ufpr.br/~paulojus/embrapa/Rembrapa/Rembrapase19.html>)

$$var(P) = (1 - f) \frac{\hat{P}\hat{Q}}{n - 1} \quad (3)$$

A variância para amostra aleatória simples sem reposição é:

[1] 0.0002603261

As variâncias para cada um dos 38 estratos da amostragem estratificada , também com erro amostral de 0,03 são:

[1] 0.0002338952 0.0002323326 0.0002350701 0.0002361649 0.0002367956
 [6] 0.0002313803 0.0002366867 0.0002367229 0.0002338952 0.0002347298
 [11] 0.0002336587 0.0002328549 0.0002329472 0.0002354510 0.0002350183
 [16] 0.0002330019 0.0002332435 0.0002344779 0.0002336006 0.0002364097
 [21] 0.0002334794 0.0002354458 0.0002330029 0.0002340013 0.0002309827
 [26] 0.0002341772 0.0002380515 0.0002350400 0.0002361014 0.0002345042
 [31] 0.0002343996 0.0002344487 0.0002358720 0.0002368040 0.0002365114
 [36] 0.0002344265 0.0002357191 0.0002337215

Notadamente cada um dos estratos apresenta uma variância menor que o planejamento sem considerar estratos.

Agora é importante saber o quanto a amostra estratificada é "*melhor*" , ou seja, o quanto este planejamento reduz a variância em comparação ao planejamento sem estratos.

Para isso existe um conceito bastante importante que é o chamado *design effect* ou efeito do planejamento que compara a variância dos dois plano:

$$EPA = \frac{Var(\hat{P}_{est})}{Var(\hat{P}_{AAs})} = 0.9 \quad (4)$$

Portanto, com o mesmo erro amostral, consegue diminuir a variância do estimador de \hat{P} em aproximadamente 10%

3 Considerações Finais

O plano estratificado proporcional produz variâncias sempre menores que aquelas produzidas por uma amostra aleatória simples (com ou sem reposição) de mesmo tamanho, e

este ganho é maior quanto maior for a variância dos estratos, isto é, quanto maior for a diferença entre as estimativas dos parâmetros dos estratos. Este fato é importantes para o estatístico desenhar um plano amostral mais conveniente de acordo com a possibilidade de aplicação dos pesquisadores.

References

- [1] Bolfarine, H and Bussab, W. *Elementos de Amostragem*. Editora Blucher-(2005).
- [2] R Core Team (2016). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Viena, Austria. <http://www.R-project.org/>.
- [3] <http://www.censo2010.ibge.gov.br/sinopseporsetores/?nivel=st>.
- [4] <http://www.censo2010.ibge.gov.br/cnefe/>.
- [5] <http://www.leg.ufpr.br/~paulojus/embrapa/Rembrapa/Rembrapase19.html>.