

Aula 5: Amostragem e tipos de amostra

Métodos e Técnicas de Pesquisa I – 2015

Márcia Lima
Murillo Maschner

Roteiro:

- **A lógica da amostragem por survey.**
 -
- **Conceitos elementares**
- **Tipos alternativos de amostras**
- **Exemplos: pesquisas amostrais**

A lógica da amostragem

Porque usar amostras?

Estudos censitários

x

Estudos amostrais

Amostragem de desenho de pesquisa

O método de obtenção de amostras é muito relevante em determinar a robustez e a validade de resultados empíricos. Existem tantas maneiras de se realizar a definição de uma amostra, que há especialidades dentro da **Estatística** que se dedicam à investigação dos melhores métodos para se fazê-lo, sendo Amostragem e Planejamento de Experimentos as duas mais conhecidas.

A definição de procedimentos de amostragem deve, portanto, ser pensada em relação ao desenho de pesquisa que se pretende desenvolver. Os procedimentos científicos de obtenção de dados amostrais mais usualmente utilizados são de dois tipos: experimentos e estudos observacionais. Aqui, nos concentraremos no 2º tipo, mais usualmente aplicado à pesquisa na área de ciências sociais.

Amostragem

Qualquer conjunto ou amostra da população permite resolver o problema da relação entre estatísticas da população e parâmetros estimados a partir de uma amostra desta população?

A resposta é **NÃO**! E por isso existem diferentes técnicas de obtenção de amostras, adequadas a diferentes desenhos de pesquisa e que contam, cada qual, com vantagens e desvantagens para o desenvolvimento de investigações de caráter científico nas ciências humanas.

População e amostra são dois conceitos básicos que sustentam o desenvolvimento da Inferência Estatística: procuramos inferir características de uma dada população a partir da obtenção de informações a respeito de uma parcela (amostra) desta população

Conceitos e terminologias

Universo: agregação teórica ou hipotética de todos os elementos definidos num survey. Conjunto total de elementos que constituem uma área de interesse analítico. Ex: Brasileiros.

População: agregação teoricamente especificada de elementos do survey. Ex: Maiores de 16 anos; nascidos ou naturalizados brasileiros.

Conceitos e terminologias

População do survey: agregação de elementos do qual a amostra do survey é de fato selecionada. Ex: Ambiente urbano; cidades acima de 5 mil habitantes etc.

Unidade da Amostra: elementos ou conjunto de elementos considerados para seleção em alguma etapa da amostragem. Ex: Cidades, bairros, firmas, escolas etc.

Elemento: unidade sobre a qual a informação é coletada. Ex: pessoas, famílias, escolas etc.

Conceitos e terminologias

- **Moldura de amostragem:** lista de unidades da qual se retira a amostra. Em geral são as molduras de amostragem que definem a população de um survey. Ex: Registros administrativos, cartografias, listas telefônicas.
- **Unidade de análise e unidade de observação:** em geral podem ser o mesmo, mas não necessariamente.
Exemplo:
 - unidade de análise: domicílio
 - unidade de observação: pessoa de referência.

Exemplo: Perfil dos eleitores do Brasil

Universo: eleitores brasileiros

População (caracterização dos eleitores brasileiros)

- i) brasileiros e naturalizados na faixa etária de 18 a 70 anos;
- ii) brasileiros e naturalizados com idade de 16 e 17 anos e acima de 70 anos que optaram por votar.

População do survey: seleção dos **indivíduos (elementos)** que serão extraídos para composição da amostra

Desenho Amostral

Viés: relaciona-se com as características das unidades de análise (ocupação, sexo, cor etc.) A ausência de viés garante que as unidades sejam diversificadas a ponto de corresponderem a população.

Precisão: garantia de que as várias aplicações daquela desenho amostral proporcionarão resultados similares e portanto, confiáveis.

Desenho Amostral

- **Margem de erro:** informa o quão perto a estatística da amostra está em relação ao parâmetro da população.
- **O erro amostral** (distância entre a amostra e o universo que pretende representar será tanto menor quanto:
 - Maior for a amostra; e
 - Mais homogêneos os elementos de análise.
- **Nível ou intervalo de confiança:** Permite expressar a precisão das estatísticas em termos de níveis de confiança, estabelecendo que essas estatísticas estão dentro de um intervalo especificado no parâmetro.
- *Numa pesquisa onde um candidato tem 30% das intenções de voto, afirma-se que:*
 - *A pesquisa tem uma margem de erro de três pontos percentuais para mais ou para menos e 95% de confiança*
 - *OU*
 - *Com 95% de confiança a proporção dos eleitores que irão votar no candidato X encontra-se no intervalo de 30% +/- 3% (27% a 33%)*



Variáveis

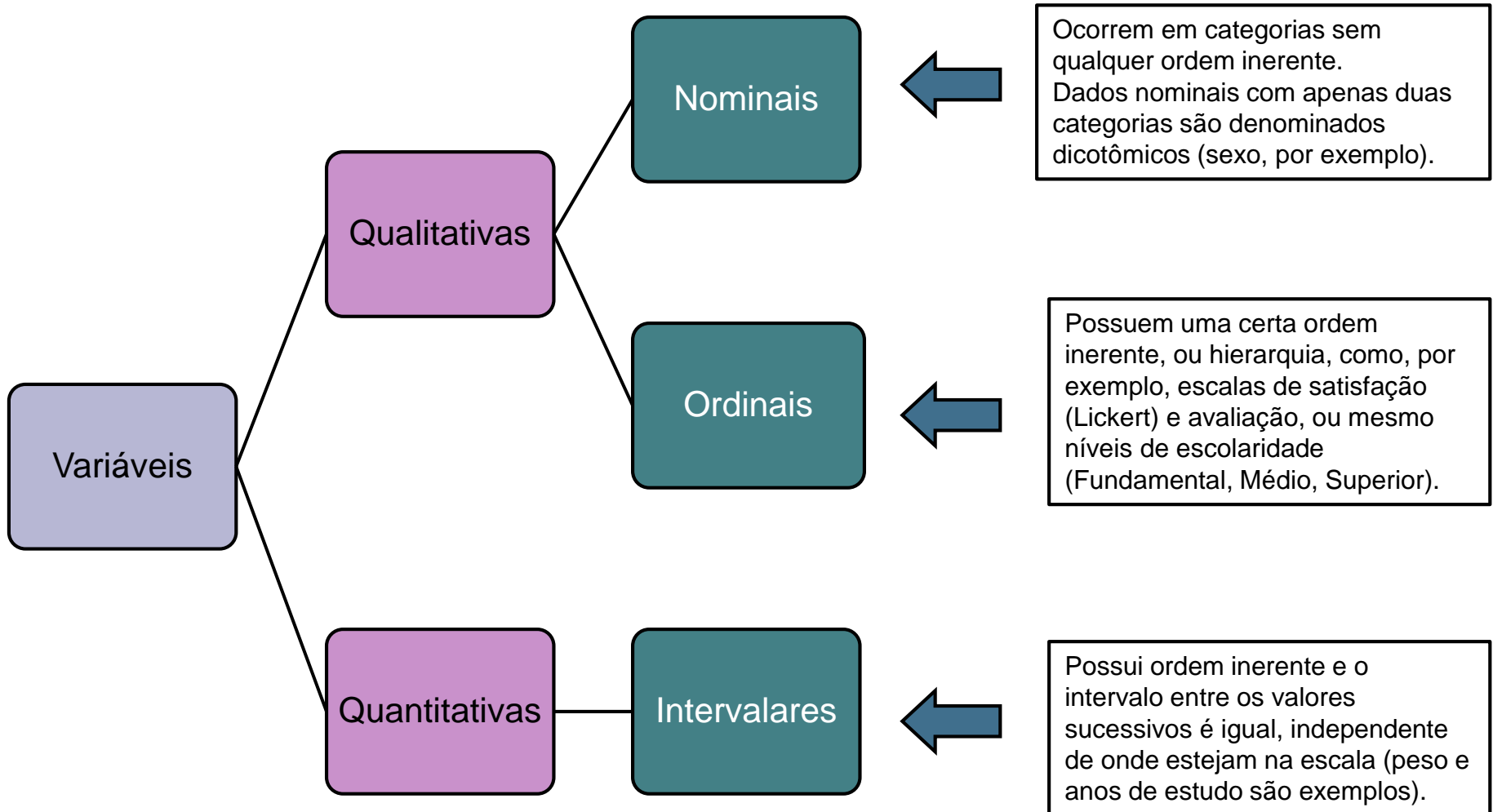
Uma variável é uma característica/atributo que pode ser observada (ou mensurada) para cada caso, a partir da definição de uma unidade de análise.

De um ponto de vista prático, podemos entender como variáveis as perguntas de um questionário estruturado, ou, em sistemas de informação, cada um dos atributos do conjunto de casos de uma dada unidade de análise (municípios, países, estabelecimentos, etc).

No nosso exemplo, temos 6 variáveis:

- Idade, renda, sexo, aparelho, plano contratado e pacote de dados

Tipos de Variáveis



Variáveis Categóricas (ou Qualitativas)

Variáveis Categóricas Ordinais

Uma variável pode ser classificada como ordinal quando a ordenação dos valores importa para a avaliação dos casos.

Exemplos

- Grau de instrução (fundamental, médio, superior)
- Escalas Lickert (totalmente satisfeito, parcialmente satisfeito, etc...)

Variáveis Categóricas Nominais

Uma variável pode ser classificada como ordinal quando não faz sentido pensar em uma ordenação hierárquica clara entre as diferentes categorias em que os casos se enquadram.

Exemplos:

- sexo
- raça
- município

Variáveis Contínuas (ou quantitativas)

Variáveis Contínuas

Uma variável deve ser classificada como contínua quando é utilizada para a mensuração de um fenômeno que possa ser expresso através de um valor numérico. Isto significa que este tipo de variável é baseado em uma medida baseadas em unidades intervalares equidistantes.

Exemplos

- Anos de escolaridade (diferente de grau de instrução!)
- Renda (domiciliar, per capita, do trabalho, etc)
- idade
- Peso / altura
- Valores monetários (R\$, US\$, €, £, etc.)

Probabilística

Toda unidade tem uma probabilidade conhecida se ser selecionada (aleatoriedade)

Todas as combinações têm igual probabilidade de aparecer

Aleatória Simples

Sistemática

Estratificada

Conglomerada

Nem todas as combinações têm igual probabilidade de aparecer

Não-Probabilística

Não se conhece a probabilidade de inclusão de cada unidade da população.

Cotas

Intencional

Por julgamento

Amostragem

As fases de um processo de amostragem

Depois de se identificar os dados que deverão ser recolhidos e o instrumento (questionário estruturado, por exemplo) a utilizar para a coleta, o passo seguinte consiste em definir um processo de amostragem adequado ao tipo de dados e ao instrumento de análise.

É necessário que se estabeleça de partida um **plano de amostragem** de acordo com a população alvo, com a definição da população pesquisada e com um processo adequado de organização da pesquisa, e portanto, de obtenção das informações junto à população.

“Dados coletados de forma descuidada podem ser tão inúteis que nenhum procedimento estatístico consegue salvá-los”

Amostragem

Tipos de Amostras

Probabilísticas

As amostras probabilísticas dependem da existência de uma listagem que defina o tamanho da população a ser investigada, informação a partir da qual se obtém um universo amostral de tamanho N .

. **Amostra aleatória simples** – Sorteio simples. Neste caso sabemos a priori que as chances de seleção de cada um dos elementos do universo é igual e diferente de zero. Podem ser feitas com e sem reposição, com procedimentos definidos para o caso da necessidade de reposições.

. **Amostra Sistemática** – Define-se aleatoriamente um primeiro elemento no universo amostra, define-se um intervalo dependendo do n^o de casos com os quais se quer compor a amostra, selecionando os casos $K+\text{intervalo}$, de maneira sistemática.

Amostragem

Tipos de Amostras

Probabilísticas

. **Amostra estratificada** – Consiste na definição de grupos relativamente homogêneos na população, mutuamente exclusivos (estratos), procedendo à amostragem aleatória simples dentro de cada estrato. Amostras estratificadas tendem a ser especialmente eficientes quando há alta heterogeneidade entre-grupos e baixa heterogeneidade intra-grupos.

. **Amostra por conglomerados ou “clusters”** – Partindo-se de uma divisão pré-definida de unidades territoriais homogêneas, seleciona-se aleatoriamente determinadas unidades territoriais, incluindo na amostra todos os elementos que compõe estas unidades. Este é um caso no qual, diferente das técnicas anteriores, basta existir uma listagem das unidades territoriais primárias, e não da totalidade da população de interesse.

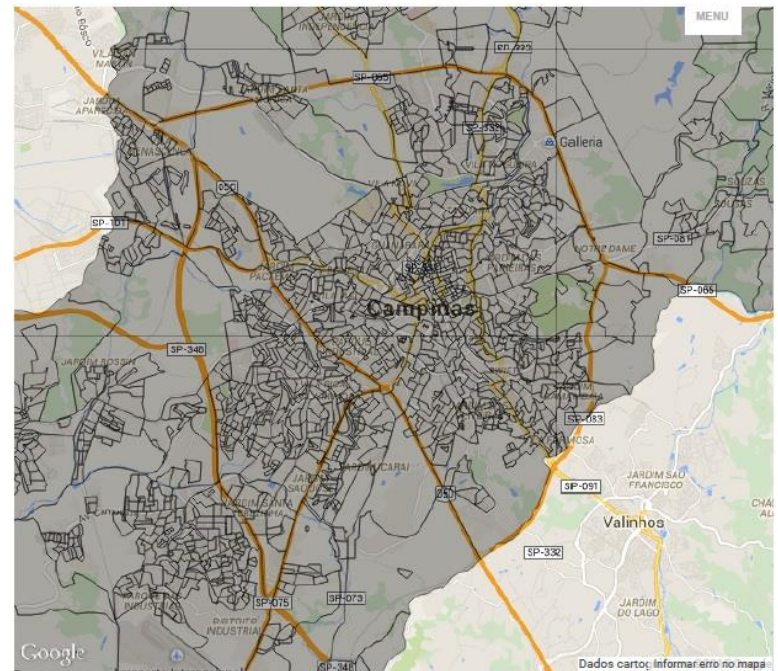
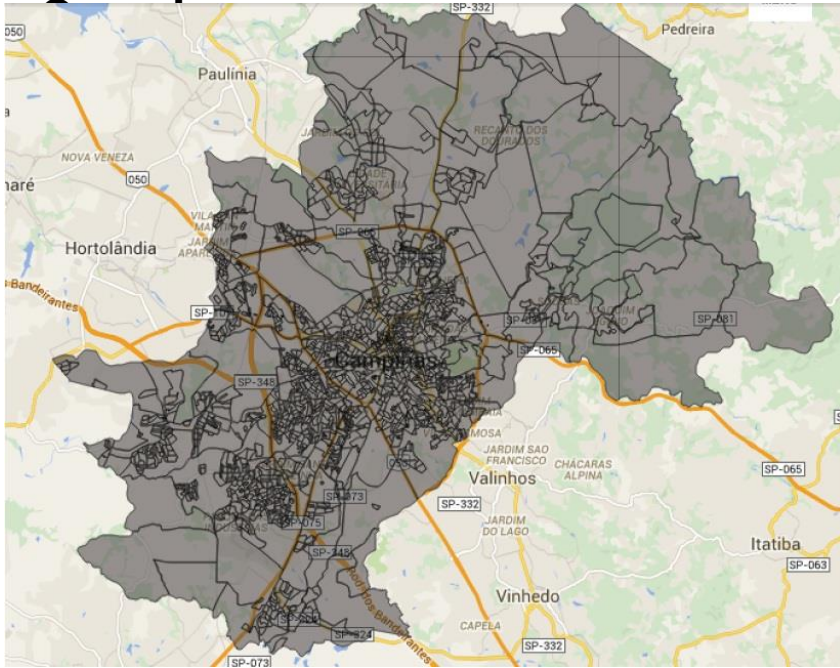
Amostragem

Na PNAD (Pesquisa Nacional por Amostragem de Domicílios) do IBGE

“trata-se de um **modelo aleatório**, em **3 estágios de seleção** (municípios, setores e domicílios), com **estratificação** das unidades de 1º estágio e com probabilidades de seleção proporcionais ao tamanho para os municípios e setores, enquanto que os domicílios serão **selecionados com probabilidades iguais** dentro dos setores”

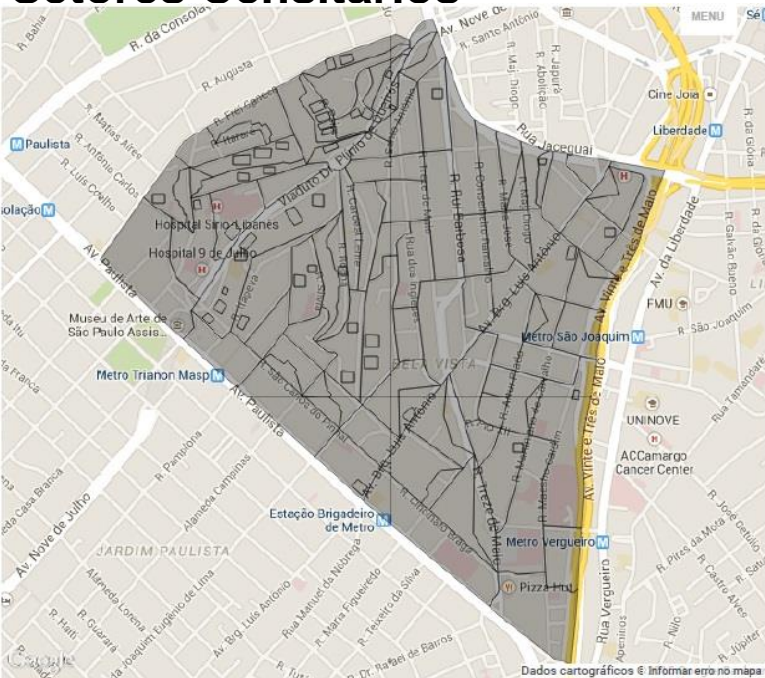
Amostragem

Setores Censitários

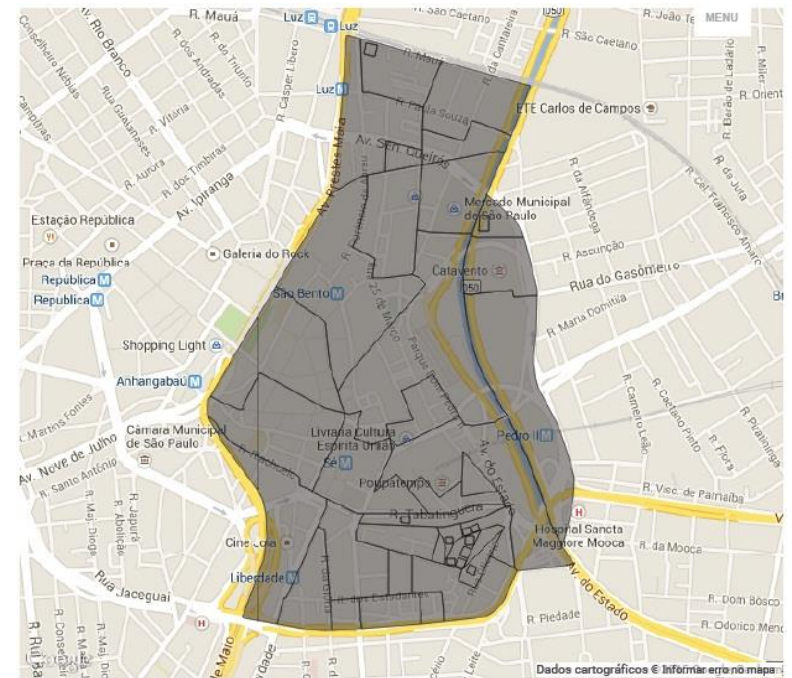


Amostragem

Setores Censitários



Bela Vista



Sé

Amostras não probabilísticas

Não é possível estabelecer as probabilidades de inclusão de cada indivíduo.

- Para sua maior eficácia, é necessário um conhecimento prévio e confiável da moldura da amostragem;
- Maior suscetibilidade ao problema de viés.

Amostragem

Tipos de Amostras

Não Probabilísticas

. **Amostra acidental ou de conveniência** – trata-se de um subconjunto da população formada pelos elementos *que se foi possível obter*. Tipicamente utiliza-se este tipo de amostragem em (a) pré-testes e (b) aproximações iniciais do problema e do objeto de pesquisa, momentos nos quais o pesquisador não tem clareza sobre as características a investigar ou sobre a adequação do instrumento de pesquisa;

. **Intencional** - Composta por elementos da população selecionados intencionalmente pelo pesquisador, por possuírem características típicas ou representativas da população.

Amostragem

Tipos de Amostras

Não Probabilísticas

- . **Bola de Neve (“snowball”)** – Tipo de amostra intencional no qual são selecionados determinados indivíduos que funcionam como “sementes” a partir dos quais são obtidas indicações de outros indivíduos pertencentes à população de interesse. Nesse caso, a amostra vai crescendo como uma bola de neve.

Amostras não probabilísticas

Amostra por cotas

Procedimentos:

1. Divide a população em subgrupos
2. Seleciona-se uma cota de indivíduos em cada categoria pertinente.
3. Define o número total de entrevistas de divide de acordo com as proporções

Exemplo: Pesquisa de opinião política e eleitoral

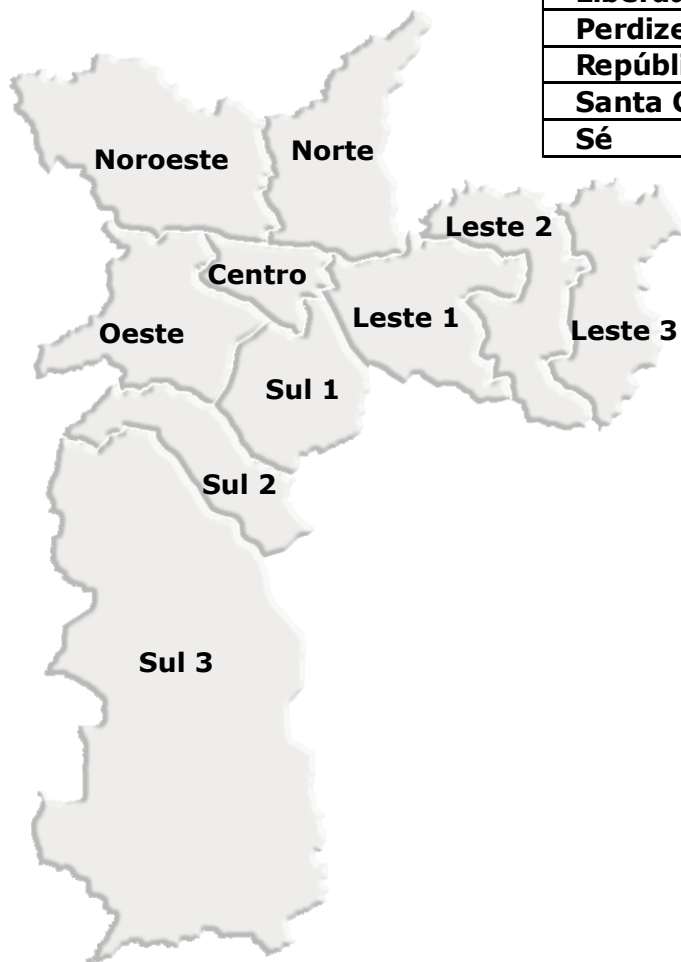
Definir variáveis relevantes. No caso da pesquisa eleitoral: sexo, idade, e município em que o eleitor vota.

2. Obter dados censitários com os números referentes às variáveis escolhidas. Em geral esses dados são do IBGE
3. Calcular as proporções das variáveis para a população. No caso das pesquisas eleitorais, os eleitores de um município, UF ou do país.
4. Definir o tamanho da amostra
5. Multiplicar as proporções de cada variável obtida ou o cruzamento destas pelo tamanho da amostra. Assim se obtém o número de entrevistas por cota.

Amostras intencional ou por julgamento

- Seleção baseada no próprio conhecimento da população, dos seus elementos e da natureza das metas da pesquisa;
- O interesse está na opinião (ação, intenção etc.) de determinados elementos da população;
- Em geral, os elementos que irão compor essa amostra são considerados como representativos de um determinado grupo.
- (Exemplo: donas de casa, pessoas ligadas a determinados grupos, usuários de drogas etc.)

Composição 10 regiões - [Total da amostra]



CENTRO (5%)	LESTE 1 (14%)	NOROESTE (12%)	SUL 1 (10%)
Pari	Móoca	Perus	Vila Mariana
Barra Funda	Penha	São Domingos	Campo Belo
Bom Retiro	Sapopemba	Jaraguá	Moema
Bela Vista	Tatuapé	Anhanguera	Ipiranga
Brás	Vila Formosa	Brasilândia	Cambuci
Consolação	Vila Matilde	Cachoeirinha	Cursino
Liberdade	Vila Prudente	Casa Verde	Jabaquara
Perdizes	Belém	Freguesia do Ó	Sacomã
República	Água Rasa	Limão	Saúde
Santa Cecília	Aricanduva	Pirituba	SUL 2 (8%)
Sé	Artur Alvim	NORTE (9%)	Campo Limpo
	Carrão	Santana	Santo Amaro
	São Lucas	Jaçanã	Vila Andrade
	LESTE 2 (11%)	Tremembé	Campo Grande
	Cangaíba	Mandaqui	Cidade Ademar
	Ermelino Matarazzo	Tucuruvi	Pedreira
	Ponte Rasa	Vila Guilherme	SUL 3 (13%)
	Cidade Líder	Vila Maria	Capão Redondo
	Itaquera	Vila Medeiros	Socorro
	Parque do Carmo	OESTE (8%)	Cidade Dutra
	São Mateus	Pinheiros	Jardim São Luís
	São Rafael	Itaim Bibi	Grajaú
	Vila Jacuí	Jaguareé	Jardim Ângela
	LESTE 3 (11%)	Jaguara	Marsilac
	Jardim Helena	Vila Leopoldina	Parelheiros
	São Miguel Paulista	Alto de Pinheiros	
	Cidade Tiradentes	Butantã	
	Guaianases	Jardim Paulista	
	Iguatemi	Lapa	
	Itaim Paulista	Morumbi	
	José Bonifácio	Raposo Tavares	
	Vila Curuçá	Rio Pequeno	
	Lajeado	Vila Sônia	

Peso eleitores por distrito

cód	distritos	peso	cód	distritos	peso	cód	distritos	peso	cód	distritos	peso
1	Centro	5,2460	2	LESTE 1	12,9370	5	NOROESTE	11,9570	8	SUL 1	10,2330
56	Pari	0,3940	53	Mooca	0,5150	61	Perus	0,9520	90	Vila Mariana	1,1590
6	Barra Funda	0,3430	59	Penha	1,1310	95	São Domingos	0,6870	15	Campo Belo	0,5870
9	Bom Retiro	0,3220	76	Sapopemba	2,4610	42	Jaraguá	1,4020	32	Moema	0,7660
7	Bela Vista	0,5440	80	Tatuapé	0,7370	3	Anhanguera	0,6150	34	Ipiranga	0,9590
10	Brás	0,2720	85	Vila Formosa	1,0090	11	Brasilândia	2,2900	14	Cambuci	0,4510
26	Consolação	0,5220	91	Vila Matilde	0,8730	13	Cachoeirinha	1,3450	27	Cursino	0,9590
49	Liberdade	0,4870	93	Vila Prudente	0,8870	21	Casa Verde	0,8660	38	Jabaquara	2,0750
60	Perdizes	0,9020	8	Belém	0,4510	29	Freguesia do Ó	1,4310	68	Sacomã	2,1320
66	República	0,4940	1	Água Rasa	0,7230	50	Limão	0,7300	77	Saúde	1,1450
69	Santa Cecília	0,6870	4	Aricanduva	0,8940	63	Pirituba	1,6390	9	SUL 2	7,2770
78	Sé	0,2790	5	Artur Alvim	1,1950	6	NORTE	8,5500	17	Campo Limpo	1,6460
			20	Carrão	0,8090	70	Santana	1,0590	71	Santo Amaro	0,6300
			72	São Lucas	1,2520	39	Jaçanã	0,9300	83	Vila Andrade	0,6220
1	CENTRO	5,2460	3	LESTE 2	11,0240	81	Tremembé	1,4950	16	Campo Grande	0,8660
2	LESTE 1	12,9370	18	Cangaíba	1,3310	51	Mandaqui	1,0950	22	CidadeAdemar	2,2180
3	LESTE 2	11,0240	28	Matarazzo	1,0660	82	Tucuruvi	0,9730	58	Pedreira	1,2950
4	LESTE 3	11,6310	64	Ponte Rasa	0,9520	86	Vila Guilherme	0,5510	10	SUL 3	13,0150
5	NOROESTE	11,9570	24	Cidade Lider	1,1090	89	Vila Maria	1,0880	19	Capão Redond	2,1390
6	NORTE	8,5500	37	Itaquera	2,0750	92	Vila Medeiros	1,3590	79	Socorro	0,4360
7	OESTE	8,1300	57	Parque do Carm	0,6370	7	OESTE	8,1300	23	Cidade Dutra	1,7030
8	SUL 1	10,2330	73	São Mateus	1,5500	62	Pinheiros	0,6010	46	Jardim São Luí	2,3040
9	SUL 2	7,2770	75	São Rafael	1,0800	35	Itaim Bibi	0,8090	30	Grajaú	2,5620
10	SUL 3	13,0150	87	Vila Jacuí	1,2240	41	Jaguareé	0,4580	43	Jardim Ângela	2,2970
		100,0000	4	LESTE 3	11,6310	40	Jaguara	0,3650	52	Marsilac	0,3790
			44	Jardim Helena	1,2520	88	Vila Leopoldina	0,3720	55	Parelheiros	1,1950
			74	São Miguel	0,9300	2	Alto de Pinheir	0,4650			
			25	Tiradentes	1,8100	12	Butantã	0,5800			
			31	Guaianases	0,8870	45	Jardim Paulista	0,8010			
			33	Iguatemi	0,8440	48	Lapa	0,5440			
			36	Itaim Paulista	2,1390	54	Morumbi	0,4440			
			47	José Bonifácio	1,1310	65	Raposo Tavares	0,8090			
			84	Vila Curuçá	1,2810	67	Rio Pequeno	1,1020			
			96	Lajeado	1,3570	94	Vila Sônia	0,7800			