UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO BACHARELADO EM SISTEMAS DE INFORMAÇÃO

Amostragem

RECIFE 2015

ANTONIO JORGE FERREIRA DELGADO FILHO EDVAN JOAQUIM SOARES JÚNIOR LUCAS FILIPE VIEIRA DA SILVA MICHEL FERREIRA DA SILVA MELO

AMOSTRAGEM

Trabalho apresentado no curso de bacharelado Da Universidade Federal Rural de Pernambuco, Curso de Sistemas de Informação.

Orientação: Profo Lucian Bejan

RECIFE 2015

SUMÁRIO

População e Amostra	4
Tipos de Amostragem	6
Amostragem Probabilística	6
Amostragem Não Probabilística	8
Exemplos de Amostragem da área do curso	10
Tamanho de uma amostra	10
Bibliografia	14

População e Amostra

Para estudarmos População e Amostra precisamos conhecer os seus conceitos. Mas afinal o que é uma População e o que é uma amostra?

Toda pesquisa realizada pela estatística precisa atender a um certo público alvo, com um certo objetivo, pois é com base nesse conjunto de pessoas que os dados são coletados e analisados de acordo com o princípio da pesquisa. Esse público alvo recebe o nome de população (em alguns casos também chamado de universo estatístico) e constitui um conjunto de pessoas que apresentam características próprias, por exemplo: os usuários de um plano de saúde, os membros de uma equipe de futebol, os funcionários de uma empresa, os eleitores de um município, estado ou país, os alunos de uma escola, os associados de um sindicato, os integrantes de uma casa e várias situações que envolvem um grupo geral de elementos. A população também pode ser relacionada a um conjunto de objetos ou informações. Cada elemento da população é conhecido como, unidade estatística. O número de elementos da população designa-se por dimensão da população e representa-se por N. Na estatística, a população pode ser classificada como finita e infinita, ou seja, a dimensão da população pode ser finita ou infinita.

População finita é chamada assim, pois nesses casos o número de elementos de um grupo não é muito grande, a entrevista e a análise das informações devem abordar a todos do grupo. Como exemplos de populações finitas temos: As condições das escolas particulares na cidade de Goiânia (se observarmos esse grupo, chegaremos à conclusão de que o número de escolas particulares em Goiânia é considerado finito, já que pode ser abordado todo o grupo), jogadores profissionais de um time, alunos de uma determinada escola e etc.

População infinita é assim conhecida, pois o número de unidades estatisticas ou elementos nesse caso é muito elevado, sendo considerado infinito. Como exemplo de uma população infinita podemos citar: população do estado de Pernambuco, os resultados obtidos (cara ou coroa) em sucessivos lançamentos de uma moeda, a temperatura em cada ponto de uma cidade, os pontos de uma reta e etc.

Há autores como BARBETTA (2001), que divide a população em dois tipos: população alvo e população acessível. População Alvo é o conjunto de elementos que queremos abranger em nosso estudo. São os elementos para os quais desejamos que as conclusões oriundas da pesquisa sejam válidas. Por exemplo: uma pesquisa feita em amostra de pequenas e médias empresas de confecção da Região Metropolitana do médio Vale do Itajaí, seria extrapolável para todas as empresas de confecção da Região Metropolitana. A população alvo, portanto, constitui-se de todas as empresas de confecção da Região Metropolitana do Vale

do Itajaí. Já a População acessível seria o conjunto de elementos que queremos ou podemos abranger por questões de racionalidade física, em nosso estudo e que são passíveis de serem observados, com respeito às características que se pretende levantar. Muitas vezes torna-se impossível observar todos os elementos de uma população. Então, em vez de pesquisar toda a população, examinamos apenas uma pequena parte que denominamos amostra. Logo, a população acessível, segundo BARBETTA (2001), será a amostra.

Mas o que viria a ser uma amostra? Amostra diz respeito a um subconjunto da população, fração ou uma parte do grupo. Em alguns casos seria impossível entrevistar todos os elementos de uma população (por exemplo uma população infinita, como a citada anteriormente no conceito), pois levaria muito tempo para concluir o trabalho ou até mesmo seria financeiramente inviável, dessa forma, o número de entrevistados corresponde a uma quantidade determinada de elementos do conjunto, uma amostra.

O objetivo da Estatística é o estudo de populações. Mas como podemos estudar essas populações? Para que isso seja possível, são realizados os censos e os processos de amostragem. Num censo são observados todos os elementos da população relativamente aos diferentes atributos que estão a ser objeto de estudo. Há diversos conceitos para censo, entre eles estão: "é o conjunto de dados estatísticos que informa diferentes características dos habitantes de uma cidade, um estado ou uma nação". Mas para nosso objetivo, definiremos censo como o levantamento efetuado sobre toda uma população.

Enquanto um censo envolve o exame de todos os elementos de uma dado grupo, a amostragem envolve o estudo apenas de uma parte dos elementos. A amostragem tem como finalidade fazer generalizações sem precisar examinar todos os elementos de um dado grupo. A amostragem é o processo de determinação de uma amostra a ser pesquisada. Existem técnicas científicas para a seleção correta de amostras. Entre essas técnicas, as mais conhecidas são a amostragem aleatória simples, a amostragem sistemática e a amostragem estratificada.

Na amostragem aleatória simples qualquer elemento da população tem a mesma probabilidade de ser escolhido. Como por exemplo, se pretendermos selecionar uma amostra de 30 alunos de uma escola de 600 alunos, numeramos de 1 a 600 os alunos da escola e, seguidamente, escolhemos ao acaso 30 desses números, que correspondem aos 30 alunos da amostra. Na amostragem sistemática os elementos da amostra são escolhidos a partir de uma regra estabelecida. Por exemplo, para selecionarmos uma amostra de 30 alunos de uma escola de 600 alunos, depois de numerados todos os alunos da escola, podemos, por exemplo, escolher um aluno de 20 em 20 a partir do primeiro aluno selecionado, que é escolhido ao acaso de entre o primeiro grupo de 20. E a amostragem estratificada utiliza-se quando a população está dividida em estratos ou grupos diferenciados.

Exemplo: Na selecção de 30 alunos de uma escola secundária, considerando cada ano de escolaridade como estrato, escolheríamos em cada um desses anos um determinado número de alunos por um dos processos anteriores.

Tipos de Amostragem

- Amostragem Probabilística

Nas sondagens e estudos de opinião usam-se amostras selecionadas por um processo de amostragem probabilística, isto é, amostras em que todos os sujeitos incluídos na grelha de amostragem têm a mesma probabilidade (diferente de zero) de serem escolhidos para virem a integrar a amostra.

As técnicas de amostragem probabilística asseguram, com uma determinada margem de erro, que estão representados na amostra (de forma proporcional à sua representação na população) todos os subgrupos relevantes que constituem a população alvo. Desta forma garante-se que os resultados obtidos com o estudo dos sujeitos da amostra podem ser generalizados, com uma determinada margem de erro, para a população alvo.

A amostragem probabilística refere-se a *procedimentos que utilizam alguma forma* de selecção aleatória dos seus membros.

Para se ter um método de selecção aleatória, é necessário definir um*procedimento* que garanta que as diferentes unidades da população têm probabilidades iguais de serem escolhidas. Não é suficiente que a escolha dos membros da amostra seja aleatória. É igualmente determinante para que uma amostragem seja probabilística que todos os membros da população tenham a mesma probabilidade de serem escolhidos.

Os procedimentos de amostragem probabilística devem cumprir 4 critérios (Chochran, 1977, p. 9):

- 1. Deve ser possível definir o conjunto de amostras distintas que o procedimento é capaz de selecionar.
- 2. Cada amostra possível tem uma probabilidade conhecida de selecção.
- 3. As amostras são selecionadas por um processo aleatório no qual cada amostra tem a mesma probabilidade de ser selecionada.
- 4. O método para calcular a estimativa deve conduzir a uma estimativa única para qualquer amostra específica.

Para a definição e selecção de uma amostra através de um processo de amostragem probabilística é necessário definir a população-alvo e dispor de uma grelha de amostragem com a identificação dos membros dessa população [nas sondagens eleitorais a população-alvo são os sujeitos que podem votar nessa eleição, isto é, os eleitores, e a grelha de amostragem utilizada é o "caderno eleitoral" (documento que identifica os eleitores). Nas sondagens relativas a eleições nacionais não é possível, em Portugal e por razões legais, dispor dos cadernos eleitorais oficiais pelo que se usa em sua substituição, frequentemente, a lista telefônica (com todos os problemas que isso coloca pelo facto de nem todos os

eleitores estarem identificados nas listas telefônicas e de as listas telefônicas incluírem vários elementos que não pertencem à população dos eleitores).

As amostragens probabilísticas podem ser: Amostragem aleatória simples; amostragem sistemática; Amostragem aleatória estratificada; Amostragem por clusters; amostragem multi-etapas.

Na amostragem aleatória simples, todos os elementos da população tem a mesma probabilidade de pertencerem à amostra. Essa amostragem pode ser sem repetição, que é quando o elemento que já foi sorteado não continua no sorteio ou com repetição, quando o elemento sorteado continua no sorteio, podendo ser escolhido novamente.

Exemplo: se tivermos uma população de 6 sujeitos [A, B, C, D, E e F] e quisermos selecionar uma amostra de 2 sujeitos, cada um destes 6 sujeitos deverá ter a mesma probabilidade de ser escolhido (1/6) e todos os subconjunto de dois elementos possíveis ([A,B], [A,C], [A,D], [A,E], [A,F], [B,C], [B,D], [B,E], [B,F], [C,D], [C,E], [C,F], [D,E], [D,F] e [E,F]) deverão ter, igualmente, a mesma probabilidade de ser escolhidos (1/15).

A **amostragem sistemática** consiste em selecionar as unidades elementares da população em intervalos pré-fixados. Para funcionar, a técnica requer a listagem prévia da população, que deverá ser homogênea e uma atenção especial na periodicidade dos dados.

Exemplo: Em uma produção diária de peças automotivas, podemos, a cada 20 peças produzidas, retirar uma para pertencer a uma amostra da produção de um dia.

A amostragem aleatória estratificada consiste em separar a população em grupos homogêneos e, então, fazer a seleção via amostragem aleatória simples dos elementos de cada grupo. Esta técnica exige um conhecimento maior da população.

Exemplo: Número de pessoas que vivem nos domicílios de Recife. Dividir os domicílios em níveis socioeconômicos e depois selecionar domicílios em cada nível aleatoriamente (Renda baixa, média, alta).

Na **amostragem por clusters**, a população é dividida fisicamente em conglomerados. Depois, selecionam-se aleatoriamente os conglomerados que farão parte da amostra, ao passo que todos os elementos dos conglomerados selecionados serão amostrados. É muito utilizada quando há necessidade de se realizar entrevistas ou observações em grandes áreas.

Exemplo: se se pretender estudar o comportamento dos doentes hospitalizados e se houver estudos anteriores que permitam concluir que o comportamento desses doentes é essencialmente o mesmo em todos os hospitais (*clusters*), então poderá fazer sentido escolher aleatoriamente só alguns hospitais (*clusters*) e estudar todos

os doentes que estão nos hospitais seleccionados, em vez de seleccionar aleatoriamente uma amostra de doentes de todos os hospitais.

Na amostragem multi-etapas seleciona-se em primeiro lugar, aleatoriamente, uma amostra por *clusters* – repare-se que é muito mais fácil obter uma lista por *clusters* (por exemplo, de escolas) do que uma lista exaustiva dos elementos que compõem a população (por exemplo, todos os alunos). Seguidamente pode-se realizar, ou não, uma segunda etapa, na qual são escolhidos aleatoriamente alguns elementos dos *clusters* selecionados na fase anterior ou, então, continuando com a seleção de *clusters* até se chegar às unidades elementares.

Exemplo: para uma sondagem sobre as eleições, dividimos o país em regiões. A seguir dividimos as regiões em pequenos grupos populacionais com dimensão semelhante e selecionamos aleatoriamente alguns destes centros. Dividimos estes grupos em freguesias e selecionamos algumas aleatoriamente. No fim, escolhemos alguns habitantes dessa freguesia para inquirir.

- Amostragem Não Probabilística

A amostragem não probabilística é totalmente subjetiva, já que se baseia nas decisões pessoais do pesquisador. Por não haver, por causa disso, uma seleção realmente aleatória dos elementos da amostra, sua variabilidade amostral não pode ser calculada com precisão. Em contra partida, ela se torna barata e mais fácil de ser aplicada.

As amostragens não probabilísticas utilizam-se em três tipos de situações:

- 1. Estudos em grupos cujos elementos são difíceis de identificar e contactar (por exemplo, membros de *gangs* juvenis);
- 2. Estudos com grupos específicos em que razões éticas impedem que se identifiquem todos os elementos desses grupos, pelo que se entrevistam apenas voluntários (por exemplo, sujeitos portadores de determinada doença);
- 3. Investigações em situações piloto (por exemplo, sujeitos que participam numa acção de formação)

As principais técnicas de amostragem não probabilística podem ser intencionais e não intencionais (ou de conveniência):

1. Amostragens intencionais

■ Amostragem Bola-de-neve – o entrevistador estabelece contacto inicial com alguns sujeitos previamente identificados como membros do grupo que se pretende estudar e estes sujeitos põem o investigador em contacto com outros membros desse grupo e assim sucessivamente (por exemplo, consumidores de drogas que indicam ao investigador outros consumidores que aceitam participar no estudo).

Exemplo: Feita a entrevista, o entrevistado poderá indicar onde encontrar outros torcedores do América e assim sucessivamente até atingir o tamanho necessário.

■ Amostragem por Quotas – Nesta situação podem utilizar-se as amostragens de conveniência ou de bola-de-neve, mas procura-se que o número de sujeitos entrevistados pertencentes a sub-grupos significativos da população (que podem ser a área geográfica, idade, sexo, estatuto socio-profissional, etc.) corresponda à percentagem desse sub-grupo na população a estudar. Por exemplo, se estivermos a estudar uma população de consumidores de droga e se pretendermos ter uma amostra de 1000 sujeitos utilizando uma amostragem em bolade-neve e se soubermos que entre a população de consumidores de droga existem 60% de sujeitos do sexo masculino e 40% do sexo feminino, entrevistam-se sujeitos de cada um destes subgrupos à medida que os vamos contactando até se atingir a percentagem que foi previamente identificada para cada um deles (ou seja, até termos uma amostra com 600 homens e 400 mulheres). No caso do nosso exemplo, isso significa que logo que se consigam entrevistar 600 sujeitos do sexo masculino já não entrevistamos mais nenhum sujeitos do sexo masculino e procuramos apenas encontrar sujeitos do sexo feminino até que se atinjam os 400 sujeitos deste grupo. Embora haja uma certa analogia entre a amostragem por quotas e a amostragem aleatória estratificada elas não devem ser confundidas. No primeiro caso (amostragem por quotas) a escolha dos sujeitos a incluir na amostra não é aleatória (vão-se entrevistando sujeitos a que se tem acesso por um processo de amostragem de conveniência ou de bola-de-neve até se atingir o valor da quota), enquanto no segundo caso a amostragem é aleatória (os membros da população são previamente identificados numa grelha de amostragem, os sujeitos a entrevistar são escolhida de forma aleatória a partir dessa grelha, só depois é que são contactados e só os sujeitos incluídos previamente na amostra é que são entrevistados).

Exemplo: Admita-se que se deseja pesquisar o "trabalho das mulheres". A primeira tarefa é descobrir uma proporção de características na população. Imagine que haja 47% de homens e 53% de mulheres na população. Logo uma amostra de 50 pessoas deverá ter 23 homens e 27 mulheres. Então o pesquisador receberá uma quota para entrevistar 27 mulheres.

■ Focus Group – Consiste na entrevista a um pequeno grupo de sujeitos-10 a 20 elementos – escolhidos pelo investigador por serem representativos de uma população particular (jovens, potenciais consumidores de um produto, membros de uma profissão, etc.

Exemplo: Os clientes de uma marca qualquer vem agora com o novo modelo focus group. Dos mesmos, criadores podem observar a discussão por detrás de uma janela espelhada de um lado. Os participantes não podem observar para fora mas os clientes podem ver a discussão. Normalmente uma câmara de vídeo é usada para gravar a discussão para ser vista por outros que não puderam estar presentes no momento. Os pesquisadores examinam mais do que apenas as palavras. Eles estão interessados nas expressões faciais, linguagem corporal e carnal do corpo e

dinâmica de grupo. Por vezes são criadas transcrições da gravação, nos pontos mais importantes.

Casos críticos –

2. Amostragens não intencionais

■ Amostragem de conveniência — entrevistam-se sujeitos a que se tem acesso imediato e directo (por exemplo, estudantes de uma turma de que o investigador é professor, doentes de um determinado serviço hospitalar,...)

Exemplos de Amostragem da área do curso - Computação quantica por amostragem de bósons

Uma das técnicas mais promissoras para a realização de computação quântica é a que envolve a chamada amostragem de bósons. Imagina-se que um sistema desse tipo seja capaz de resolver problemas que são inviáveis — ou pelo menos muito difíceis — para computadores clássicos.

Entretanto, uma das dificuldades inerentes à técnica é que chega a ser difícil até mesmo validar o funcionamento de um dispositivo de amostragem de bósons, simulando sua operação e colhendo os resultados relevantes do arranjo.

Um avanço importante nesse sentido acaba de ser obtido por um grupo ítalobrasileiro encabeçado por Nicolò Spagnolo, da Universidade de Roma. Em um esforço experimental envolvendo interferometria e interferência multifótons, os cientistas demonstraram que é possível verificar o bom funcionamento de um dispositivo por amostragem de bósons.

"Nossos resultados pavimentam o caminho para experimentos de amostragem de bósons maiores cujo funcionamento, apesar de ser não-trivial para simular, podem ser certificados contra hipóteses alternativas", escrevem os pesquisadores.

Tamanho de uma amostra

É muito importante que o pesquisador saiba avaliar a relevância clínica da diferença estatisticamente significante pois raras são as vezes em que o pesquisador pode trabalhar com todos os elementos da população, visto que este trabalha com tempo, energia e recursos econômicos limitados. Assim, faz-se necessário um planejamento do tamanho da amostra a ser estudada para se evitar altos custos com amostras exageradamente grandes, que podem tornar diferenças irrelevantes em estatisticamente significativas.

Como já foi dito, esse processo de escolha dos indivíduos que estarão contidos na amostra é denominado amostragem, e é através dela que o pesquisador pode generalizar conclusões, estendendo-as para toda toda a população da qual essa amostra foi retirada.

Uma amostra não representa perfeitamente uma população, quanto a isso não há dúvidas. Isso acaba levando a uma margem de erro que denominamos erro

amostral, este representa a diferença entre um resultado amostral e o verdadeiro resultado populacional.Não podemos evitar a ocorrência do erro amostral, mas através de uma escolha adequada do tamanho de uma amostra nós podemos limitar o valor do erro. Obviamente, quanto maior for o tamanho da amostra menor sera o erro cometido.

A capacidade de uma de seguir uma distribuição estatística acaba determinando sua classificação como grande ou pequena. As grandes amostras são aquelas a densidade de probabilidade de forma definida, seguindo melhor as funções as funções de distribuição adotadas, o que não é possível verificar em pequenas amostras

Em uma grande amostra não há unanimidade no número de eventos que a define. A norma ASHRAE 41.5-75 (1975) indica 20 eventos (n > 20), enquanto que Triola (1998) já indica 30 eventos (n > 20). Isso acontece devido ao Isso se deve ao comportamento dos níveis de confiança associados ao valor do desvio padrao σ em função do número de graus de liberdade ν , quando o número de eventos é inferior ao limite citado acima esse comportamento passa a variar de forma importante. Podemos observar tal comportamento através da figura

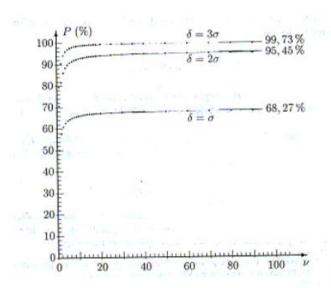


Figura 1- Níveis de confiança P em função dos graus de liberdade v (Fonte: Vuolo, 1998)

O valor médio \underline{x} é a melhor estimativa da média populacional μ , ou valor verdadeiro, em uma grande amostra. Ao valor da média é associado um intervalo de confiança que obedece a uma dada probabilidade.

As equações aplicadas para grandes e pequenas amostras são:

média aritmética (1)

$$\overline{x} = \frac{\sum x_i}{n}$$

desvio padrão para grandes amostras (2)

$$\sigma = \left(\frac{\sum (x_i - \overline{x})^2}{n}\right)^{1/2}$$

desvio padrão experimental ou convencional (3)

$$s = \left(\frac{\sum (x_i - \overline{x})^2}{n - 1}\right)^{1/2}$$

Determinamos o tamanho de uma amostra grande através do cálculo da incerteza da média

$$u_{\bar{x}} = \frac{k \sigma}{\sqrt{n}}$$

Dela podemos isolar o valor do número de eventos n, tal que

$$n = \left(\frac{k\,\sigma}{u_{\bar{x}}}\right)^2$$

Pequenas amostras - distribuição t de Student

Em amostras com número de eventos inferior a 30 ou 20, o valor do desvio padrão não é mais conhecido estatisticamente onde o número de eventos do denominador n-1 é conhecido por graus de liberdade v. Podemos compreender a subtração de um evento numa pequena amostra pelo fato que a média é empregada para o cálculo de grandezas estatísticas, e portanto está compreendida.

Não estamos mais tratando com uma distribuição gaussiana, pelo fato do desvio padrão não ser reconhecido. Assim, a melhor opção que se adapta neste caso é a distribuição t de Student, desenvolvida por William Gosset (1876-1937) que trabalhava para uma cervejaria Guinness. Essa distribuição varia conforme o número de eventos, tem uma forma simétrica, e aproxima-se da distribuição de Gauss para v>30.

Podemos calcular a média de uma pequena amostra através de

$$u_{\bar{x}} = \frac{t \, s}{\sqrt{n}}$$

onde t

Onde t é o valor da distribuição para uma dada confiabilidade e um número de graus de liberdade v, s o desvio para um número de graus de liberdade v, e n é o número total de eventos da amostra.

Tabela de valores de t-Studant para diferentes níveis de confiabilidade

ν	Nível de confiabilidade			
	68,27%	95,45%	99,73%	
1	1,84	13,97	235,80	
2	1,32	4,53	19,21	
3	1,20	3,31	9,22	
4	1,14	2,87	6,62	
5	1,11	2,65	5,51	
6	1,09	2,52	4,90	
7	1,08	2,43	4,53	
8	1,07	2,37	4,28	
9	1,06	2,32	4,09	
10	1,05	2,28	3,96	
15	1,03	2,18	3,59	
20	1,03	2,13	3,42	
25	1,02	2,11	3,33	
30	1,02	2,09	3,27	
40	1,01	2,06	3,20	
50	1,01	2,05	3,16	
∞	1,00	2,00	3,00	

Observe que ao se estimar a incerteza de medição, na realidade o que se faz é estimar o desvio padrão da população a partir do desvio padrão da amostra, que subestima o primeiro. O valor estimado, portanto, é o que se deve usar em futuras medições, com um número infinito de graus de liberdade.

Bibliografia

http://www.brasilescola.com/matematica/populacao-amostras.htm

http://www.inf.furb.br/siic/sidw/apresentacao_teses/azambuja/populacaoeamostra.ph

http://www.ipb.pt/~cmca/estatistica1

http://www.pucrs.br/famat/viali/graduacao/matematica/material/laminaspi/TI_1.pdf

http://www.significados.com.br/censo/

http://www.cavalcanteassociados.com.br/utd/UpToDate168.pdf

http://www.geste.mecanica.ufrgs.br/medterm/tamanho_amostra.pdf

http://www.fesppr.br/~centropesq/Calculo_do_tamanho_da_amostra/Tamanho%20d a%20Amostra%20-%201.pdf

http://www.eecis.udel.edu/~portnoi/classroom/prob_estatistica/2006_1/lecture_slides/aula17.pdf

https://sondagenseestudosdeopiniao.wordpress.com/amostragem/amostras-probabilisticas-e-nao-probabilisticas/