



ESTATÍSTICA ORIENTADA À CIÊNCIA DE DADOS



População e Amostra



População

A população refere-se ao conjunto completo de elementos ou indivíduos que compartilham características específicas de interesse em um estudo.



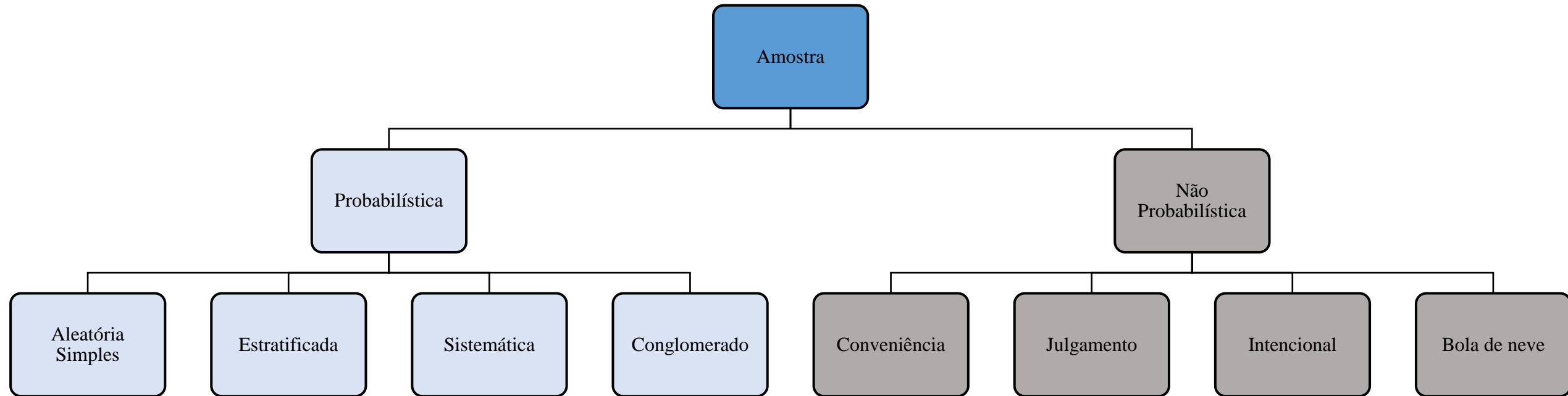
Amostra

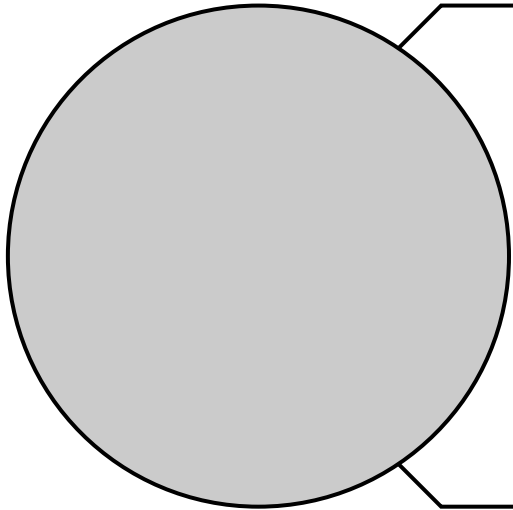
A amostra é um subconjunto representativo selecionado da população, usado para fazer inferências e generalizações sobre a população como um todo;

A escolha de uma amostra adequada é crucial para garantir que os resultados obtidos sejam confiáveis e aplicáveis à população de interesse.

Característica	População	Amostra
Definição	Conjunto completo de elementos ou indivíduos de interesse no estudo	Subconjunto selecionado da população para análise
Objetivo	Avaliar características da totalidade dos elementos	Inferir características da população a partir de um grupo representativo
Tamanho	Geralmente grande, incluindo todos os elementos	Menor do que a população, mas representativo dela
Seleção	Não é selecionada; é o conjunto total de interesse	Selecionada usando técnicas de amostragem

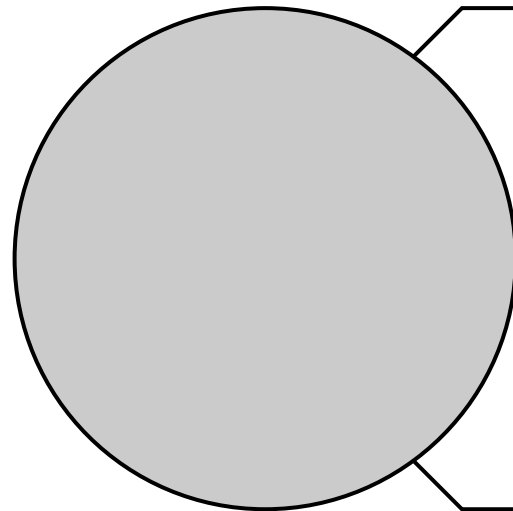
Característica	População	Amostra
Importância	Fornece informações sobre o universo de interesse	Permite generalizações sobre a população
Viés	Não está sujeita a viés de seleção	Pode conter viés se a amostra não for representativa
Exemplo	Todos os alunos de uma escola formam a população	Uma amostra aleatória de alunos é selecionada para uma pesquisa





Probabilística

Envolve a seleção de elementos da população de forma aleatória e com probabilidade conhecida



Não Probabilística

A escolha dos elementos é feita de forma não aleatória, com base em critérios subjetivos ou disponibilidade

Amostragem probabilísticas

Tipo	Definição	Vantagens	Desvantagens
Aleatória simples	Cada elemento da população tem a mesma probabilidade de ser selecionado para a amostra.	Garante representatividade; Fácil de implementar; Cálculo do erro amostral.	Exige lista completa da população; Inválida para populações amplas ou dispersas.
Estratificada	Divide a população em subgrupos (estratos) e seleciona amostras aleatórias em cada um.	Representa heterogeneidade; Reduz erro amostral; Garante inclusão de todos estratos.	Requer informações prévias sobre a população.
Sistemática	Seleciona elementos em intervalos regulares após escolher aleatoriamente o primeiro.	Mais eficiente que a AAS; Fácil de implementar em populações ordenadas.	Sensível a padrões periódicos na população.
Por conglomerados	Divide a população em grupos; Seleciona aleatoriamente alguns grupos e inclui todos.	Eficiente para populações dispersas; Facilita logística de amostragem.	Potencialmente menos representativa; Complexidade na análise dos conglomerados.

Tipo	Aleatória Simples	Estratificada	Sistemática	Por Conglomerados
Exemplo	Em uma escola com 500 alunos, utilizar um gerador de números aleatórios para selecionar 50 alunos para uma pesquisa acadêmica.	Em uma pesquisa eleitoral, dividir a população em grupos por faixa etária (18-29, 30-49, 50+ anos) e selecionar aleatoriamente nas faixas.	Em um supermercado com 100 produtos, selecionar a cada 5º produto na prateleira após sortear o número inicial.	Em uma pesquisa nacional de saúde, selecionar aleatoriamente algumas cidades e incluir todos os indivíduos nessas cidades na amostra.

Amostragem não probabilísticas

Tipo	Definição	Vantagens	Desvantagens
Por Conveniência	Elementos são selecionados com base na disponibilidade, facilidade de acesso ou conveniência do pesquisador.	Rápida e econômica; Fácil de implementar; Adequada para estudos exploratórios.	Amostra pode não ser representativa da população; Pode conter vieses e não refletir a diversidade do grupo estudado.
Julgamento (Quota)	O pesquisador seleciona manualmente elementos da amostra com base em critérios subjetivos ou cotas pré-definidas.	Fácil de implementar; Permite representar subgrupos específicos da população.	Suscetível a vieses e subjetividade na escolha dos elementos; Não garante representatividade da população em geral.

Tipo	Definição	Vantagens	Desvantagens
Intencional (Propósito)	O pesquisador seleciona deliberadamente elementos que considera mais relevantes ou representativos.	Útil para estudos qualitativos ou quando a representatividade não é a principal preocupação.	Não garante representatividade; Pode não ser adequada para estudos que requerem inferências estatísticas generalizáveis.
Bola de Neve	Usada quando a população é difícil de alcançar ou identificar; os participantes indicam outros participantes.	Útil para grupos de difícil acesso; Ajuda a identificar participantes ocultos ou pouco conhecidos; Pode obter informações valiosas.	Não garante representatividade; Pode levar a amostras pequenas e limitadas; Suscetível a vieses e falta de diversidade.

Tipo	Por Conveniência	Julgamento (Quota)	Intencional (Propósito)	Bola de Neve
Exemplo	Entrevistar alunos de uma universidade que estão próximos ao local da pesquisa para obter respostas rapidamente.	Entrevistar 50 homens e 50 mulheres para representar igualmente as duas categorias de gênero na amostra de uma pesquisa.	Entrevistar especialistas em determinado campo para obter opiniões qualificadas sobre um tema específico de estudo.	Em um estudo sobre uma comunidade marginalizada, o pesquisador entrevista alguns membros que indicam outros membros para entrevista.

Tipos de População

População Finita x População Infinita

Aspecto	População Finita	População Infinita
Definição	Tem um número finito de elementos.	Tem um número infinito de elementos.
Amostra	O tamanho da amostra é limitado pelo tamanho da população.	O tamanho da amostra não é limitado pelo tamanho da população.
Tamanho da Amostra	Pode ser uma porcentagem significativa da população.	Geralmente é uma fração pequena da população.
Impacto na Precisão	Tamanhos de amostra maiores tendem a aumentar a precisão das estimativas.	Tamanhos de amostra menores podem oferecer estimativas precisas, mas podem ter maior variabilidade.
Generalização dos Resultados	Estimativas e conclusões geralmente se aplicam apenas à população específica.	Estimativas e conclusões têm maior potencial de generalização para uma população maior.

População Finita x População Infinita

Aspecto	População Finita	População Infinita
Cálculo do Tamanho da Amostra	Considera o tamanho total da população.	Considera a taxa de amostragem e o nível de confiança desejado.
Amostragem Aleatória Simples	Pode ser mais viável devido ao tamanho limitado da população.	Ainda é viável, mas pode ser mais difícil devido à falta de limitação no tamanho da população.
Custo da Amostragem	Pode ser mais acessível devido ao tamanho limitado da população.	Pode ser mais caro devido à necessidade de selecionar uma amostra representativa de um grande universo.
Precisão das Estimativas	Pode ser alcançada com amostras relativamente pequenas, dada a limitação do universo.	Pode exigir amostras maiores para alcançar o mesmo nível de precisão devido à variabilidade inerente.

População Finita x População Infinita

Exemplo de População Finita:

Escola com 500 alunos. Neste caso, a população é finita, pois o número total de alunos é limitado e definido (500). Qualquer amostra retirada da escola será uma amostra de uma população finita.

População Finita x População Infinita

Exemplo de População Infinita:

Todas as moedas de um determinado país. Neste caso, a população é infinita, pois o número de moedas é virtualmente incontável e está sempre sujeito a mudanças devido à produção, circulação e perda de moedas. Mesmo que seja selecionado uma grande amostra de moedas, ainda não será possível abranger todas as moedas existentes.



Tamanho da Amostra



População	Fórmula para Tamanho da Amostra	Variáveis
Infinita	$n = \frac{Z^2 * p * (1 - p))}{e^2}$	n: Tamanho da amostra Z: Valor crítico da distribuição normal p: Probabilidade estimada do evento e: Erro amostral desejado
Finita	$n = \frac{N * Z^2 * p * (1 - p)}{(N * e^2 + Z^2 * p * (1 - p))}$	n: Tamanho da amostra N: Tamanho da População Z: Valor crítico da distribuição normal p: Probabilidade estimada do evento e: Erro amostral desejado

Parâmetro	Definição	Variáveis
Valor Crítico (Z)	Valor da distribuição normal padrão associado ao nível de confiança desejado.	Z=1,96 para 95% de confiança Z=1,64 para 90% de confiança Z=2,58 para 99% de confiança
Probabilidade Estimada (p)	Proporção de sucesso (evento de interesse) na população.	0 a 1
Erro Máximo Desejado (E)	A margem de erro permitida na estimativa do parâmetro populacional.	De 0 em diante

Valor Crítico (Z)

- Utilizado para determinar os limites nos quais os resultados da amostra podem ser considerados estatisticamente significativos ou para estimar intervalos de confiança para parâmetros populacionais;
- Na distribuição normal padrão (também conhecida como distribuição z), os valores críticos correspondem aos pontos específicos ao longo da curva de distribuição em que uma determinada porcentagem da área total está localizada. Esses valores são usados para delinear regiões de interesse que ajudam a tomar decisões estatísticas;

Valor Crítico (Z)

- Os valores críticos indicam os pontos nos quais você pode estar razoavelmente confiante de que a porcentagem correspondente da área sob a curva normal está contida dentro desse intervalo.

Probabilidade Estimada (p)

- Parâmetro utilizado em cálculos de tamanho de amostra quando se está trabalhando com amostragem probabilística, como a amostragem aleatória simples. Representa a proporção de sucesso ou o evento de interesse na população;
- Por exemplo, em uma pesquisa sobre a preferência de um produto entre os consumidores, caso deseja-se estimar quantos entrevistados serão necessários, a probabilidade estimada (p) seria a proporção de consumidores que você espera que tenham a preferência pelo produto.

Probabilidade Estimada (p)

- No entanto, em muitos casos, pode-se não saber a probabilidade estimada com precisão antes de realizar a amostragem. Nesse cenário, uma abordagem comum é usar uma probabilidade estimada conservadora, geralmente 0,5 (ou 50%), que é a estimativa mais conservadora quando não há informações prévias sobre a proporção de sucesso na população;
- Essa estimativa é usada para calcular o tamanho da amostra necessário para garantir que a estimativa seja confiável, independentemente da proporção real na população.

Erro Máximo Desejado (e)

- Representa a magnitude máxima pela qual a estimativa obtida a partir da amostra pode diferir do valor real do parâmetro na população;
- O erro máximo desejado está diretamente relacionado com a precisão da estimativa. Quanto menor for o erro máximo desejado, mais precisa será a estimativa. No entanto, a redução do erro máximo desejado também pode exigir um aumento no tamanho da amostra, o que pode resultar em custos mais altos (tempo, recursos, dinheiro) para coletar e analisar os dados;

Erro Máximo Desejado (e)

- O erro máximo desejado é um valor que você define com base na precisão que deseja para a sua estimativa. Quanto menor o erro máximo desejado, maior será a precisão requerida para a estimativa;
- Por exemplo, se você estiver conduzindo uma pesquisa de opinião e deseja estimar a proporção de pessoas que apoiam uma determinada política com um erro máximo de 3%, isso significa que você está disposto a aceitar que a estimativa pode diferir do valor real em até 3 pontos percentuais.

Prática