

brazilian

Data Scientist - Curso

Manoel Teles

16 de maio de 2018

Estatística - 1

Amostragem

Conceitos

População: Alvo do estudo.

Censo: Pesquisa com toda a população, pode ser caro ou impossível inferir sobre toda a população.

Enviesamento: Você subestima ou superestima o parâmetro da população.
causas: Pesquisa de pessoas próximas ou de fácil acesso, pesquisas pela internet, sem uso de mecanismo de seleção aleatório.

Amostra - Parte de uma população (Subconjunto da população), selecionada usando alguma técnica que de chances iguais a todos os elementos da população de serem selecionados.

Uma amostra feita corretamente deve representar as mesmas características da população de onde foi retirada.

Se ela não representa a população, dizemos que ela é enviesada.

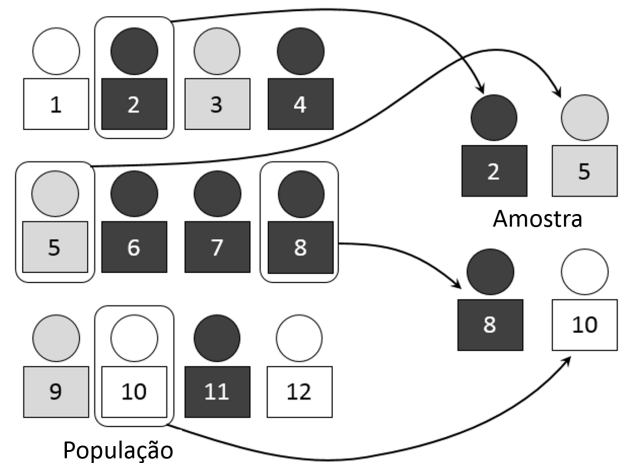
”Custo” da Amostra

Margem de erro e nível de confiança.

Variação: amostrar diferentes podem apresentar resultados diferentes.
podemos ”medir” a variação esperada.

Principais tipos de amostras

Aleatória simples - Um determinado número de elementos é retirado da população de forma aleatória. Todos os elementos da população alvo do processo de amostragem, devem ter as mesmas chances de serem selecionado para fazer parte da amostra.



Há duas formas de trabalhar com amostra aleatória simples, com reposição e sem reposição.

Com reposição: Quer dizer que uma vez que o elemento da população é selecionado, ele volta a fazer parte da população e ele passa a ter as mesmas chances de ser selecionado novamente, exemplo (Exame de doping em atletas das olimpíadas).



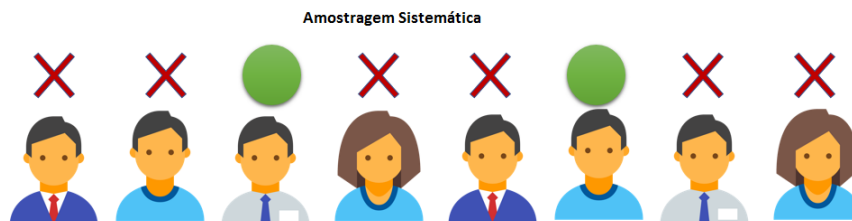
Sem reposição: Uma vez que o elemento da população é selecionado, ele não faz mais parte da população e não tem mais chance de ser selecionado novamente, exemplo (Pesquisa de intenção de votos).



Estratificada - As vezes as populações estão divididas nos chamados estratos, características comuns que os elementos tem, podem ser relacionados a raça, escolaridade ou religião.



Sistemática - Neste tipo de amostragem, é escolhido um elemento aleatório, e a partir daí, a cada N elementos um novo membro é escolhido.



Por Unidade Monetária - Neste tipo de amostra é informada uma coluna de ordenação, e uma coluna numérica que será utilizada para gerar os cálculos para produzir a amostra.

Amostragem por Unidade Monetária				
		Cientes	Débito	Cumulativo
		Austin College Austin College	R\$ 701,00	R\$ 701,00
Total de Registros:	50	Algonquin College	R\$ 1.200,00	R\$ 1.901,00
		Beijing Foreign Studies BFSU	R\$ 453,00	R\$ 2.354,00
Total de Débitos:	R\$ 108.465,00	Blackboard Inc.	R\$ 290,00	R\$ 2.644,00
		Boise State	R\$ 1.340,00	R\$ 3.984,00
Calcula o intervalo da amostra:		Bowling Green State	R\$ 245,00	R\$ 4.229,00
	Valor Total / registros = 2169	Blue Dots Consultancy Services B	R\$ 780,00	R\$ 5.009,00
		Brigham Young	R\$ 1.456,00	R\$ 6.465,00
Ordena por Clientes		British Columbia Institute of Tech	R\$ 820,00	R\$ 7.285,00
		Campus Management	R\$ 1.432,00	R\$ 8.717,00
Seleciona um número aleatório entre 1 e 2169: 1800 (Gerado no R)		CAESY Education Systems	R\$ 2.344,00	R\$ 11.061,00
		Carnegie Mellon Carnegie Mellon	R\$ 1.029,00	R\$ 12.090,00