Analytics & Inteligência Artificial

Tema da aula
Amostragem

LABDATA | FIA FUNDAÇÃO INSTITUTO DE ADMINISTRAÇÃO BUSINESS SCHOOL

Coordenadores:
Profª Drª Alessandra de Ávila Montini
Profª Dr. Adolpho Walter Pimazoni Canton

FIA Business School

NOSSOS DIFERENCIAIS | QUEM SOMOS

BUSINESS SCHOOL
Graduação, pós-graduação, MBA, Pós- MBA, Mestrado Profissional, Curso In *Company* e EAD

CONSULTING
Consultoria personalizada que oferece soluções baseada em seu problema de negócio

RESEARCH
Atualização dos conhecimentos e do material didático oferecidos nas atividades de ensino

FUNDAÇÃO INSTITUTO DE ADMINISTRAÇÃO

Líder em Educação Executiva, referência de ensino nos cursos de graduação, pós-graduação e MBA, tendo excelência nos programas de educação. Uma das principais escolas de negócio do mundo, possuindo convênios internacionais com Universidades nos EUA, Europa e Ásia. +8.000 projetos de consultorias em organizações públicas e privadas.

NOTA MÁXIMA ENADE 2019 MELHOR ESCOLA DO BRASIL
Único curso de graduação em administração a receber as notas máximas

QS WORLD UNIVERSITY RANKINGS
A primeira escola brasileira a ser finalista da maior competição de MBA do mundo

JOHN MOLSON SCHOOL OF BUSINESS
Única *Business School* brasileira a figurar no *ranking* LATAM

UN GLOBAL COMPACT
Signatária do Pacto Global da ONU

ANAMBA ASSOCIAÇÃO NACIONAL DE MBA
Membro fundador da ANAMBA - Associação Nacional MBAs

AMBA ACCREDITED
Credenciada pela AMBA - *Association of MBAs*

Executive MBA Council
Credenciada ao *Executive MBA Council*

AACSB Business Education Alliance Member
Filiada a AACSB - *Association to Advance Collegiate Schools of Business*

EFMD
Filiada a EFMD - *European Foundation for Management Development*

ESTADÃO GUIA DO MBA SELECIONADO
Referência em cursos de MBA nas principais mídias de circulação

LABDATA FIA

NOSSOS DIFERENCIAIS | QUEM SOMOS

O **Laboratório de Análise de Dados** – LABDATA é um Centro de Excelência que atua nas áreas de ensino, pesquisa e consultoria em análise de informação utilizando técnicas de ***Big Data***, ***Analytics*** e **Inteligência Artificial**.

Profª Drª Alessandra Montini

O LABDATA é um dos pioneiros no lançamento dos cursos de *Big Data* e *Analytics* no Brasil

Os diretores foram professores de grandes especialistas do mercado

+10 anos de atuação

+1000 alunos formados

Docentes

- Sólida formação acadêmica: doutores e mestres em sua maioria
- Larga experiência de mercado na resolução de *cases*
- Participação em Congressos Nacionais e Internacionais
- Professor assistente que acompanha o aluno durante todo o curso

Estrutura

- 100% das aulas realizadas em laboratórios
- Computadores para uso individual durante as aulas
- 5 laboratórios de alta qualidade (investimento +R$2MM)
- 2 Unidades próximas a estação de metrô (com estacionamento)

CONTEÚDO PROGRAMÁTICO

ANALYTICS

MACHINE LEARNING

DEEP LEARNING

ANÁLISE EXPLORATÓRIA

TÉCNICAS DE PROJEÇÃO

TÉCNICAS DE CLASSIFICAÇÃO

TÉCNICAS DE SEGMENTAÇÃO

TÉCNICAS DE ANALYTICS

LINGUAGEM

PYTHON

R

PROJETO

PROJETO ANALYTICS

Conteúdo da Aula

1. Introdução
   i. População
   ii. Censo
   iii. Amostra
2. Esquemas amostrais
   i. Amostra Aleatória Simples (AAS)
   ii. Amostra Estratificada (AE)
   iii. Amostra Sistemática (AS)
3. Cálculo do tamanho amostral & Intervalo de Confiança
   i. Média
   ii. Proporção
   iii. Média: Amostra Estratificada
4. Exercícios

1. Introdução

Introdução

O conhecimento passado sobre um determinado assunto ajuda muito nas tomadas de decisão.

Estudar levantamentos semelhantes, seja no passado, em grupos similares ou regiões diferentes são uma das melhores fontes para obter-se sugestões de como um problema pode ser resolvido.

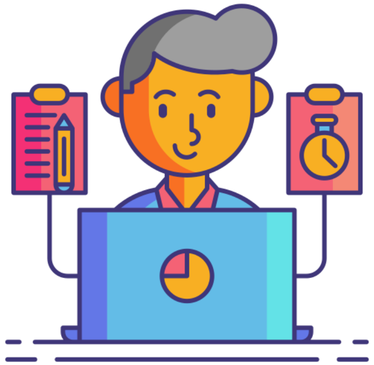
A amostragem é um procedimento realizado para extrair uma amostra de uma população, com o objetivo de inferir sobre os parâmetros populacionais.

Além de um bom planejamento do estudo e definição do esquema amostral, é crucial determinar o tamanho ótimo amostral para uma correta generalização das medidas investigadas dentro dos limites aceitáveis de incerteza.

Procedimento de obtenção dos dados

Para obtenção dos dados de uma população, dois procedimentos podem ser adotados:

1. **Censo:** coleta de dados de <u>todos os elementos</u> da população de interesse.
   - ✓ Alto custo operacional.
   - ✓ É utilizado apenas quando os dados são facilmente obtidos ou ocorre a coleta de dados para populações pequenas.
2. **Amostragem:** seleção de <u>parte dos elementos</u> da população de interesse.
   - ✓ Frequentemente utilizada para populações muito grandes ou infinitas.
   - ✓ Os dados são obtidos com um custo mais baixo e com maior velocidade.



Conceitos: População e Amostra

1.i. POPULAÇÃO | AMOSTRAGEM

- **População:** todas as observações do universo de referência.
- O tamanho da **população** será denotado por **N**.

- **Amostra:** parte de uma população de referência.
- O tamanho da **amostra** será denotado por **n**.

Todos os brasileiros: → Pesquisa eleitoral para presidência da República no Brasil.

Todos os carros produzidos em uma montadora: → Controle de qualidade de veículos para avaliar defeitos na produção.

Todos os clientes de um banco: → Treino de algoritmos para identificar padrão de cancelamento de conta corrente.

População: Infinita e Finita

- **População Finita:** pode ser quantificada numericamente:

Quantidade de alunos de uma escola.

Quantidade de carros produzidos em uma montadora.

Quantidade de clientes de um banco.

- **População Infinita:** são populações muito grandes ou de um processo contínuo em que os elementos da população são gerados indefinidamente:

Quantidade de pinguins no mundo.

Quantidade de peixes em um lago.

Quantidade habitantes do planeta.

Censo

O Censo é um estudo estatístico referente à várias informações de uma determinada população, em que os dados de todos os elementos dessa população são coletados.

No Brasil, esse estudo é realizado pelo IBGE, de dez em dez anos, e é a única pesquisa que visita todos os domicílios brasileiros.

A palavra **Censo** vem do latim “*census*” e é geralmente traduzida como “levantamento, registro, estimativa”, e tem o significado de "conjunto de dados estatísticos dos habitantes de uma cidade, província, estado ou nação".

Originou-se na Roma Antiga onde era realizado o **censo** para mapear os proprietários de terras e determinar o pagamento dos impostos.

Por que preciso de uma amostra?

Em muitos problemas do mundo real deseja-se estudar alguma variável de uma população, mas por algum motivo não é possível ter aceso a toda a população. Neste caso, pode-se analisar as informações da amostra e concluir para a população.

Teste de durabilidade da lâmpada:

Quantas lâmpadas eu preciso testar para determinar o tempo até ela queimar?

Teste de lançamento de remédio:

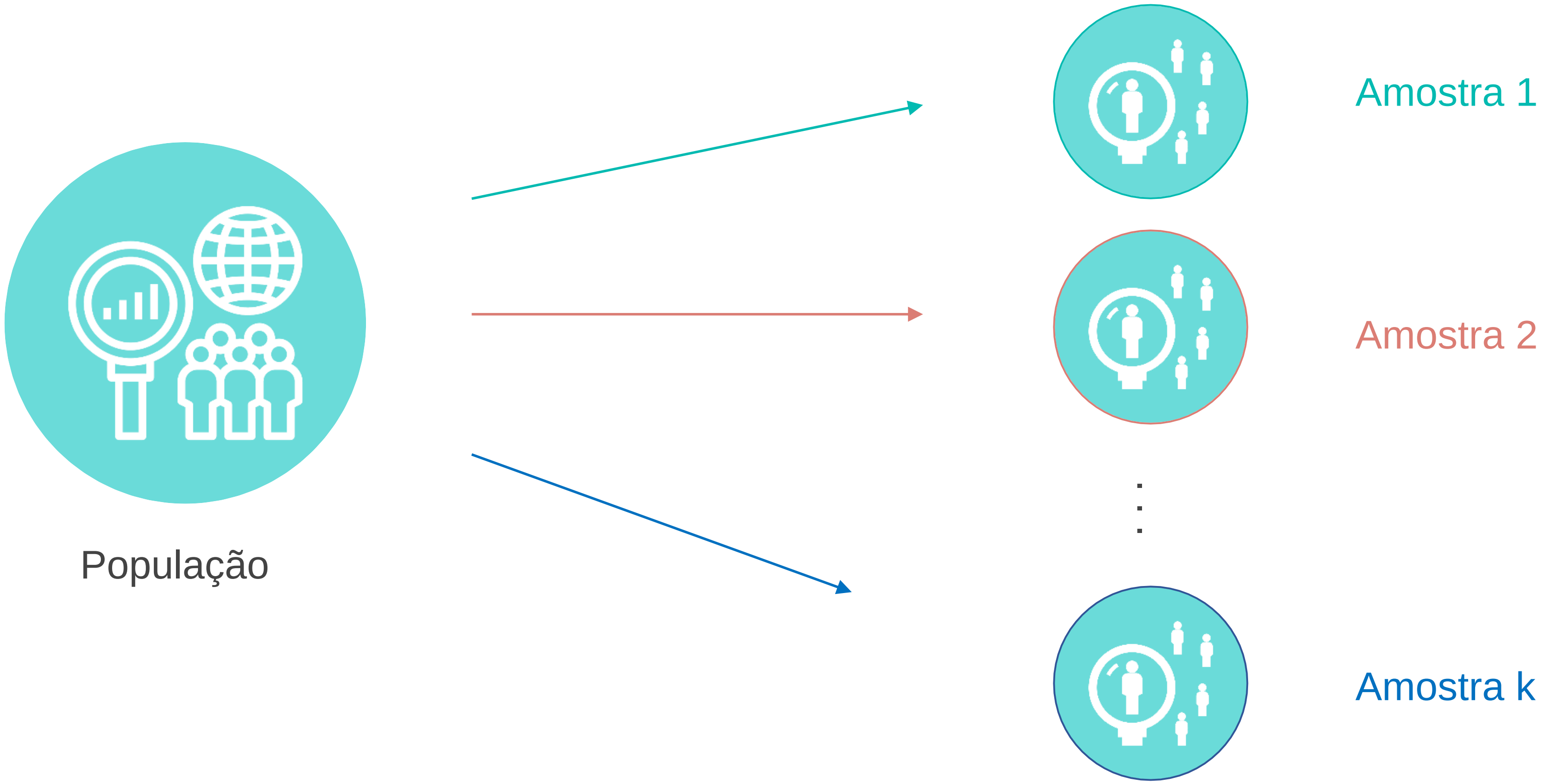
De quantos indivíduos eu preciso para concluir que o tratamento é eficaz?

O **estudo de uma amostra** possibilita a obtenção de informações a respeito de **parâmetros populacionais** por meio da observação de apenas uma parte desta população.

Processo de amostragem

1.iii. AMOSTRA | AMOSTRAGEM

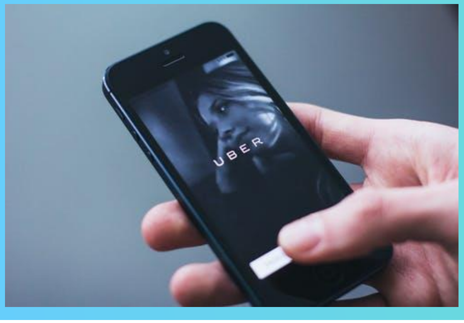
Amostragem é o processo de obtenção de amostras de uma determinada população.



Aplicações: Estudos utilizando amostra

1.iii. AMOSTRA | AMOSTRAGEM

Mobilidade urbana



Pesquisar a satisfação de seus clientes em relação ao uso do aplicativo.

Financeiro

Avaliar se houve aumento no montante investido depois de uma ação de incentivo.

Saúde

Avaliar o efeito de um novo medicamento em pacientes.

Varejo

Avaliar o aumento de ticket médio depois do envio de cupom promocional.

Telecom

Avaliar o % de clientes insatisfeitos com o serviço de telefonia.

2. Esquemas Amostrais

Conceito: Amostra Aleatória Simples

A amostragem aleatória simples é o esquema amostral mais utilizado e mais conhecido.

Uma **amostra aleatória simples (AAS)** é uma amostra de tamanho **n** elementos, extraída de uma população de **N** elementos, em que os elementos são selecionados **ao acaso**.

- Este esquema amostral é fácil de ser aplicado quando temos acesso a toda a população, por exemplo, tabulados em um banco de dados.
- Os *softwares* computacionais, geralmente, possuem a função de sorteio aleatório, que facilita a operacionalização e garantia do 'acaso'.

Em uma AAS, todos os indivíduos possuem a mesma chance de serem selecionados. A seleção ocorre ao acaso (aleatoriamente).

Exemplo: Bingo

Em um BINGO, considere a amostra oriunda do sorteio de **uma** bola dentre **N** bolas.

A bola sorteada é uma AAS, pois todas as bolas possuem a mesma probabilidade de serem sorteadas.

Aplicação: Pesquisa de satisfação

Um hospital de referência gostaria de fazer uma pesquisa sobre satisfação de atendimento com os pacientes que realizaram consulta no último mês. Esta pesquisa tem por objetivo atribuir uma nota pelo atendimento de 0-10.

Para isso, selecionou-se uma amostra de 100 pacientes aleatoriamente de um total de 5.000 que passaram por atendimento no último mês.

Todos os atendimentos são cadastrados no sistema e no sorteio via *software*, cada paciente possui a mesma probabilidade de ser sorteado para realização da pesquisa.

Aplicação: Estudo de perfil de base de dados

O diretor de uma empresa de telecomunicações gostaria de receber insights em relação ao faturamento das linhas telefônicas de seus clientes Pessoas Físicas. A carteira possui aproximadamente 70 milhões de linhas telefônicas. É necessário realizar o estudo na base populacional?

Seria possível realizar um estudo com uma amostra e inferir para base total?

Para isso, selecionou-se uma amostra de 5.000 linhas telefônicas e realizou-se um estudo sobre esta base de dados amostral. Na AAS, cada linha telefônica possui a mesma probabilidade de ser sorteada.

Conceito: Amostra Estratificada

A amostragem aleatória estratificada é um esquema amostral muito utilizado na prática.

Uma **amostra estratificada (AE)** é um esquema amostral em que é realizado uma AAS em cada grupo (ou estrato) de uma determinada população.

- A estratificação é útil para melhoria na **precisão das estimativas**, produzindo estimativas para população geral e por estrato.

Em uma AE, todos os indivíduos são separados por estrato, e dentro de cada estrato eles possuem a mesma chance de serem selecionados. A seleção dentro do estrato ocorre ao acaso (aleatoriamente).

Aplicação: Renda de clientes de cartão de crédito

Um emissor de cartão de crédito deseja estimar a renda dos clientes Pessoas Físicas. Os clientes possuem tipos de cartão diferentes (básico PF, especial, *black*). Sabendo-se da heterogeneidade da carteira em relação ao gasto com o cartão, selecionou-se aleatoriamente por estrato:

- 1.000 clientes do cartão **básico PF**, de um total de 2 milhões de clientes.
- 1.500 clientes do cartão **especial**, de um total de 500 mil cliente.
- 1.200 clientes do cartão ***black***, de um total de 50 mil clientes

Dado o estrato, todos os clientes possuem a mesma probabilidade de serem sorteados.

Aplicação: Pesquisa de conectividade na internet

Um grupo de pesquisas gostaria de investigar a quantidade de horas conectadas na internet por faixa etária. Para isso, gostaria de investigar este comportamento de forma geral e por grupo de idade: adolescentes, jovens adultos, adultos maduros e idosos.

A amostra, por estrato, foi definida segundo a representatividade por faixa etária da população na região investigada. Desta forma, a amostra deve ser retirada na seguinte proporção:

- 20% adolescentes,
- 30% jovens adultos,
- 35% adultos maduros,
- 15% idosos.

Conceito: Amostra Sistemática

A amostragem sistemática é utilizada quando a ordem dos indivíduos é relevante para o esquema amostral.

A **amostragem sistemática (AS)** é um esquema amostral em que os elementos são selecionados a partir de um critério que é aplicado a toda a população de **forma sistemática**.

Exemplo: prontuários médicos, domicílios em uma rua, entrada de clientes em um banco, etc.

- Este esquema amostral é fácil de ser aplicado e independe de termos acesso a toda a população.

Pode-se, por exemplo, selecionar um elemento a cada dois da população (k=3):

Quando se deseja extrair uma amostra sistemática de tamanho **n** de uma população de tamanho **N**, deve-se amostrar um elemento a cada **k = N/n** elementos da população.

Exemplo: Amostra Sistemática

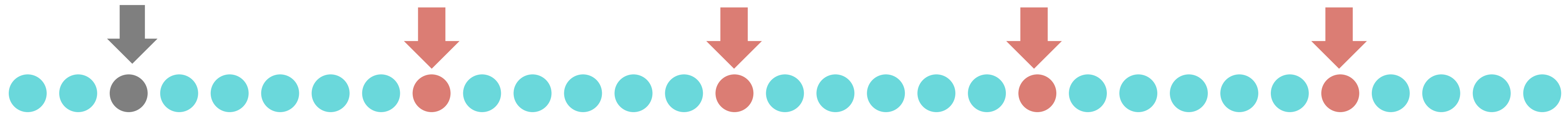
Deseja-se obter uma amostra sistemática de **n=2.000** clientes de uma base de dados de **N=12.000** clientes de uma empresa. Deve-se selecionar sistematicamente de 6 em 6 clientes.

$$K=\frac{N}{n}=\frac{12.000}{2.000}=6$$

Passos:

- Deve-se selecionar aleatoriamente um elemento dos primeiros k primeiros elementos da lista populacional.
- Suponha que entre os 6 primeiros elementos, o elemento selecionado aleatoriamente foi o terceiro.
- De forma sistemática deve-se selecionar os elementos de 6 em 6 a partir do terceiro.

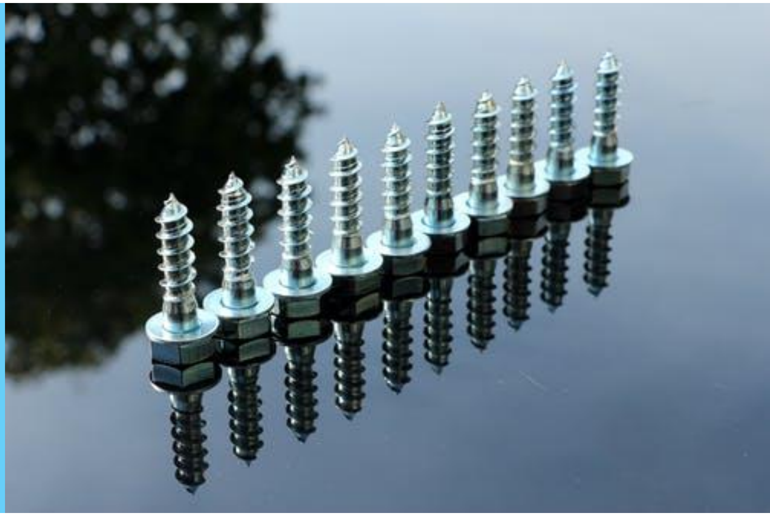
Sorteado Aleatoriamente
entre os 6 primeiros



Exemplo: Indústria de parafusos

Uma fábrica de parafusos realiza a amostragem de seus produtos para controle de qualidade.

Seleciona 1 parafuso a cada 1.000 produzidos, e o supervisor de área envia para gerência de produção a quantidade de parafusos defeituosos no final do dia. Caso seja detectado um número maior de parafusos defeituosos, a produção deve ser parada e máquina de parafusos deve receber reparos.



Aplicação: Eficácia de ações de *Marketing*

Uma *fintech* que fornece serviços 100% digitais por meio de *app* gostaria de avaliar o efeito do "kit de boas-vindas" após a solicitação do cartão de crédito. O kit é composto por um cartão de crédito com arte personalizada e kit de limpeza de celular com o logo da marca e o nome do cliente. Esta ação é cara, mas o time de *marketing* acredita que pode encantar e rentabilizar o cliente. Desta forma, a empresa, gostaria de separar um grupo de clientes que receberá o kit (grupo teste) para comparar com os demais que não receberam o kit.

A cada 100 novos cartões aprovados, 1 é selecionado para o grupo teste. A seleção da amostra ocorrerá pelo período de 1 mês, e após 6 meses de relacionamento será avaliado o indicador gasto no cartão de crédito.

3. Cálculo do tamanho amostral & Intervalo de Confiança

Como obtemos o tamanho amostral?

Dado o esquema amostral discutido no tópico anterior, agora temos o interesse em calcular o tamanho amostral ideal de tal forma que a quantidade de indivíduos selecionados seja representativa da população de interesse.

Mas.... qual o **tamanho adequado**?

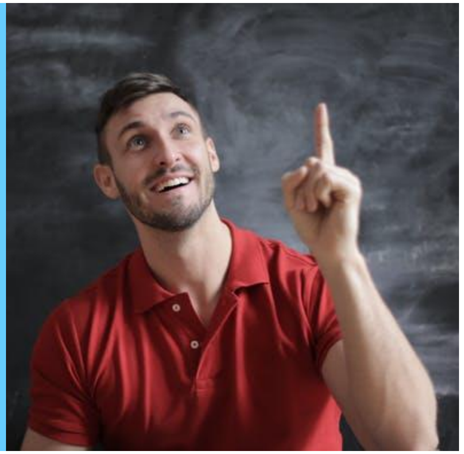
- Deve ser **grande o suficiente** para encontrar diferenças investigadas no estudo.
- Deve ser **pequena suficiente** para evitar desperdício de recurso (tempo, dinheiro ou colocar em risco algum grupo do estudo).

Para determinar o tamanho amostral é necessário definir uma **medida de incerteza** que queremos adotar no estudo, consequentemente influenciando no tamanho da amostra:

"Quanto mais certeza, mais indivíduos serão necessários para a amostra."

Intervalo de confiança γ : $[\bar{x} \pm \underbrace{\text{margem de erro}}]$

ME é o valor adicionado e subtraído da estimativa pontual para obter o intervalo com γ de confiança.



De forma bem geral, podemos obter o **tamanho amostral** da seguinte forma:

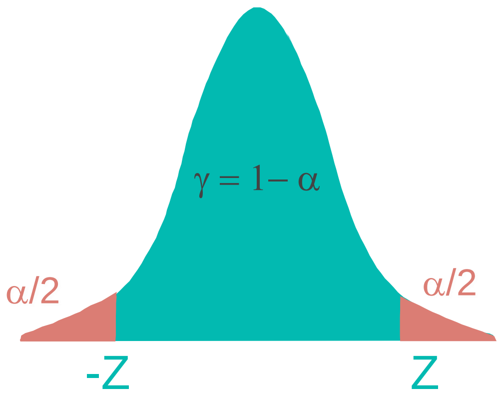
$$\mathrm{ME} = Z\frac{\sigma}{\sqrt{n}} \longrightarrow \boxed{n = Z^2\frac{\sigma^2}{\mathrm{ME}^2}}$$

Observação: Adotaremos Z para $Z_{\alpha/2}$ apenas para deixar a notação menos carregada.

Quais são as informações necessárias?

IMPORTANTE: Utilizaremos o **resultado assintótico** para os parâmetros de interesse, seja μ ou **p**, assumindo a distribuição probabilística **Normal Padrão N(0,1)** e **desvio padrão (S) a ser estimado por meio de uma amostra piloto.**

Intervalo de confiança γ : $[\bar{x} \pm \text{margem de erro}]$



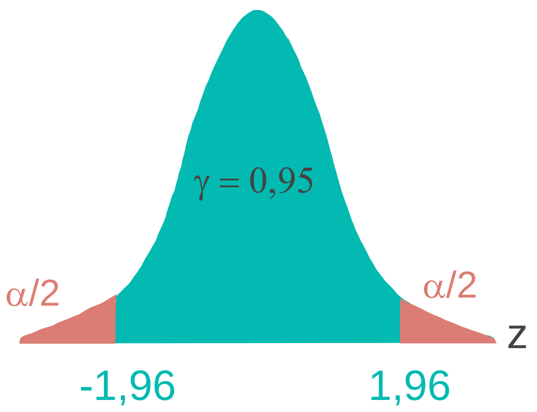
$$Z \frac{S}{\sqrt{n}}$$

Valor crítico z (**Normal Padrão**) para cálculo da margem de erro, com confiança de γ.

Para o cálculo do n amostral

Para a obtenção do valor de **n** é necessário definir a **confiança** a ser utilizada e a **margem de erro:**

1. **Confiança (γ):** geralmente considera-se 90% (Z = 1,645) ou 95 % (Z = 1,960) de confiança pela distribuição Normal Padrão.
2. **Margem de erro (ME):** é a medida de incerteza que queremos adotar no estudo.



3.i. Média

Case: Gasto com educação

Deseja-se estimar o valor da mensalidade escolar para crianças do ensino infantil em um bairro de classe média da cidade de São Paulo. Neste caso, a população é composta por todas as crianças de 3-6 anos de idade da rede privada, representada por N=15.000 crianças.

Seja μ o valor médio gasto por criança na rede privada do ensino infantil, que é denominado **parâmetro populacional**.

O valor μ não é conhecido e precisa ser estimado.

Estratégia: Retira-se uma amostra de tamanho **n** da população e estima-se o valor de μ.

Case: Gasto com educação

Exemplo de margem de erro

Considere a afirmação:

A pesquisa indicou que o valor médio pago nas mensalidades escolares é **R$1.200** com uma margem de erro de **R$100** e confiança de 95%.

Neste caso, o valor médio pago nas mensalidades escolares para região investigada estaria entre **R$1.100** e **R$1.300**, **com 95% de confiança.**

Esta afirmação quer dizer que se a amostra fosse retirada uma infinidade de vezes, 95% dos intervalos de confiança conteriam o verdadeiro valor da média μ populacional.

Margem de Erro e Tamanho amostral: média μ

3.i. MÉDIA | CÁLCULO DO TAMANHO AMOSTRAL

POPULAÇÃO INFINITA

Margem de Erro:

$$Z.\left(\frac{S}{\sqrt{n}}\right)$$

Tamanho amostral n:

$$n = \frac{Z^2.S^2}{(ME)^2}$$

POPULAÇÃO FINITA

Margem de Erro:

$$Z.\left(\frac{S}{\sqrt{n}}\right).\sqrt{\frac{N-n}{N-1}}$$

Tamanho amostral n:

$$n = \frac{Z^2.S^2.N}{(ME)^2.(N-1)+Z^2.S^2}$$

Em que :

- **ME** é a margem de erro do estudo.
- **Z** é o valor da Distribuição Normal que fornece a confiança (γ) desejada no ponto $Z_{\alpha/2}$.
- **S** é o desvio padrão estimado por uma amostra piloto.
- **N** é o tamanho populacional.

Case: Gasto com educação

Inicialmente, deve-se extrair uma amostra piloto para obter o valor de S. Suponha que neste exemplo foi extraída uma amostra piloto com desvio padrão **S=R$500** e 95% de confiança.

O **tamanho da amostra** é dado por:

$$n = \frac{Z^2 . S^2 . N}{(ME)^2 . (N-1) + Z^2 . S^2}$$

$$n = \frac{(1{,}96)^2(250.000)(15.000)}{(100)^2.(15.000-1)+(1{,}96)^2(250.000)} = 95$$

Do total de 15.000 alunos, considerando uma margem de erro de R$100 e confiança de 95%, o tamanho amostral estimado é de 95 alunos.

Case: Gasto com educação

3.i. MÉDIA | CÁLCULO DO TAMANHO AMOSTRAL

A pesquisa foi realizada com $n = 95$ pais de crianças e obteve-se uma média amostral do valor pago na mensalidade escolar de R$1.200.

O valor R$**1.200** é a estimativa pontual da mensalidade escolar das crianças do ensino fundamental do bairro em questão.

Exemplo de margem de erro:

Considere a afirmação:

A pesquisa indicou que o valor médio pago nas mensalidades escolares é **R$1.200** com uma margem de erro de **R$100**, e confiança de 95%.

Neste caso, o valor médio pago nas mensalidades escolares estaria entre **R$1.100** e **R$ 1.300**, **com 95% de confiança.**

margem de erro = $Z.\left(\frac{S}{\sqrt{n}}\right).\sqrt{\frac{N-n}{N-1}}$

Exercício: Pesquisa de satisfação

Um hospital de referência gostaria de fazer uma pesquisa sobre satisfação de atendimento com os pacientes que realizaram consulta. Esta pesquisa tem por objetivo, identificar os pontos de melhoria nos serviços de atendimento, desde recepção, exames laboratoriais, tempo de espera, atenção do médico, recomendações clínicas, etc. O paciente atribui uma nota entre 0 a10 para hospital.

Calcule o tamanho amostral considerando um total de 500 pacientes, com 95% de confiança, margem de erro de 1 ponto e S=3,5 pontos na nota da pesquisa.

O diretor de uma empresa de telecomunicações gostaria de entender o valor gasto mensalmente no serviço de telefonia pré-pago. A carteira tem aproximadamente 50 milhões de clientes.

Calcule o tamanho amostral necessário para entender o valor gasto com seus serviços, considerando o resultado da última pesquisa, com **desvio padrão amostral 20 reais. Adote 95% de confiança e margem de erro de 2 reais**.

Simulação: valores amostrais

3.i. MÉDIA | CÁLCULO DO TAMANHO AMOSTRAL

À medida que o tamanho populacional cresce, o tamanho amostral converge e os resultados dos cálculos de amostras para **população finita e infinita ficam similares**.

			População Finita	População Infinita
Z	1,96	N	n	n
S	1500	10	7,9	34,6
S2	2250000	50	20,7	34,6
ME	500	100	25,9	34,6
		1000	33,5	34,6
		5000	34,3	34,6
		10000	34,5	34,6
		50000	34,5512	34,6
		500000	34,5721	34,6

			População Finita	População Infinita
Z	1,96	N	n	n
S	5000	10	9,1	96,0
S2	25000000	50	33,1	96,0
ME	1000	100	49,2	96,0
		1000	87,7	96,0
		5000	94,2	96,0
		10000	95,1	96,0
		50000	95,8578	96,0
		500000	96,0217	96,0

3.iii. Proporção

Case: Pesquisa eleitoral

3.ii. PROPORÇÃO | CÁLCULO DO TAMANHO AMOSTRAL

Deseja-se estimar a proporção de eleitores que possuem intenção em votar no candidato João para ser representante dos funcionários de uma montadora de carros, no segundo turno das eleições.

Qual o tamanho amostral necessário para realização da pesquisa da fábrica que tem 2.000 funcionários?

Seja **p** a proporção populacional de eleitores que possuem intenção em votar no candidato João.

Como **p** é um valor relacionado a toda a população ele é denominado **parâmetro populacional**.

Case: Pesquisa eleitoral

3.ii. PROPORÇÃO | CÁLCULO DO TAMANHO AMOSTRAL

No exemplo, o valor de **p** não é conhecido e precisa ser estimado.

Estratégia: Retira-se uma amostra de tamanho **n** da população para estimar o valor de **p**.

Para estimar o valor de **p** deve ser calculada a proporção amostral de pessoas com intenção em votar no candidato João, denominada $\hat{p}$.

Para estimar a proporção amostral de pessoas com intenção em votar no candidato João, é necessário a extração de uma amostra aleatória simples de tamanho **n**.

Margem de Erro e Tamanho amostral: proporção p

POPULAÇÃO INFINITA

Margem de Erro:

$$Z.\left(\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right)$$

Tamanho amostral n:

$$n = \frac{Z^2.\hat{p}(1-\hat{p})}{(ME)^2}$$

POPULAÇÃO FINITA

Margem de Erro:

$$Z.\left(\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right).\sqrt{\frac{N-n}{N-1}}$$

Tamanho amostral n:

$$n = \frac{Z^2.\hat{p}(1-\hat{p}).N}{(ME)^2.(N-1)+Z^2.\hat{p}(1-\hat{p})}$$

Em que :

- **ME** é a margem de erro do estudo.
- **Z** é o valor da Distribuição Normal que fornece a confiança desejada.
- $\hat{p}(1-\hat{p})$ é a variância estimada que é função de $\hat{p}$ fornecido pela amostra piloto.
- **N** é o tamanho populacional.

Uma pesquisa preliminar indicou que a proporção amostral de pessoas com intenção de votar no candidato João é 0,5. Se for adotado uma margem de erro de **0,03**, o intervalo de confiança seria:

$$[0,5 \pm 0,03] = [0,47; 0,53]$$

Case: Pesquisa eleitoral

Para estimar a proporção amostral de pessoas com intenção em votar no candidato João, o tamanho da amostra necessário seria dado por, com 95% de confiança:

$$n = \frac{Z^2 . \hat{p}(1 - \hat{p}) . N}{(ME)^2 . (N - 1) + Z^2 . \hat{p}(1 - \hat{p})}$$

$$n = \frac{(1{,}96)^2 . 0{,}25 . (2.000)}{(0{,}03)^2 . (2.000 - 1) + (1{,}96)^2 . 0{,}25} = 696{,}07 = 696.$$

A amostra necessária para estimar a proporção de pessoas com intenção de votar no candidato João, com 95% de confiança, seria de **696 funcionários** da empresa.

Case: Pesquisa eleitoral

3.ii. PROPORÇÃO | INTERVALO DE CONFIANÇA

Selecionou-se aleatoriamente uma amostra de **696** funcionários da fábrica, e com base nesta amostra a intenção em votar no candidato João foi de 0,54, com uma margem de erro de 0,03.

O valor **0,54** é a **estimativa pontual** da verdadeira proporção de pessoas com intenção de votar no candidato João.

A estimativa intervalar para a verdadeira proporção de pessoas com intenção em votar no candidato João, considerando 95 % de confiança, é dada por:

[0,54 ± 0,03].

Desta forma, a proporção populacional de pessoas com intenção em votar no candidato João pode estar entre **0,51 e 0,57**, com 95 % de confiança.

Conclusão: com 95% de confiança, o candidato João deve ganhar as eleições para representante dos funcionários da empresa.

O RH de uma empresa estima que 20% dos seus funcionários estão insatisfeitos com seus líderes diretos, e se tivessem oportunidade mudariam de gestor. Desta forma, o RH gostaria de confirmar esta suspeita e entrevistar presencialmente os funcionários para entender mais a fundo o motivo da insatisfação.

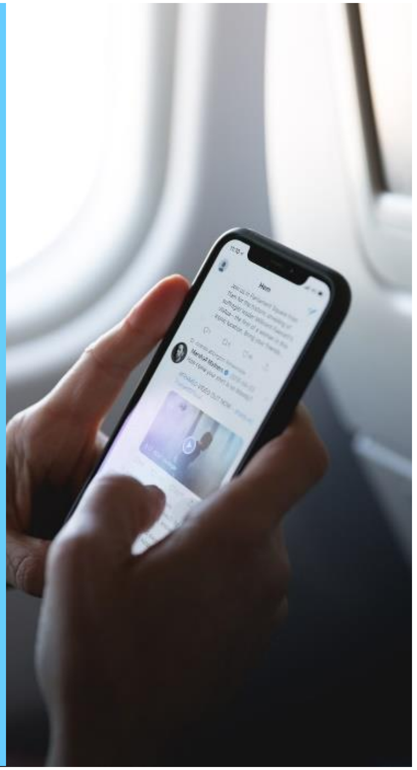
Qual a quantidade de funcionários que o RH da empresa precisaria entrevistar presencialmente, dentre o total de 1.000 funcionários, adotando uma margem de erro de 0,025 e confiança de 95%?

Case: Campanha de CRM em Telecom

O diretor de *marketing* de uma empresa de telecomunicações gostaria de avaliar uma parceria com uma empresa de entretenimento, oferecendo a seus clientes “ingressos de cinema com 50% de desconto”. Seu objetivo com esta ação é criar um vínculo maior com a marca e consequentemente fidelizar o cliente. Considere como indicador de fidelização o % de cancelamento.

Para avaliar a relevância da parceria, primeiramente ele gostaria de conduzir um estudo em uma amostra, oferecendo o benefício apenas para uma pequena parcela de seus clientes, e caso realmente o benefício for efetivo, fecharia a parceria com a empresa de entretenimento e ofereceria o benefício para todos os seus clientes.

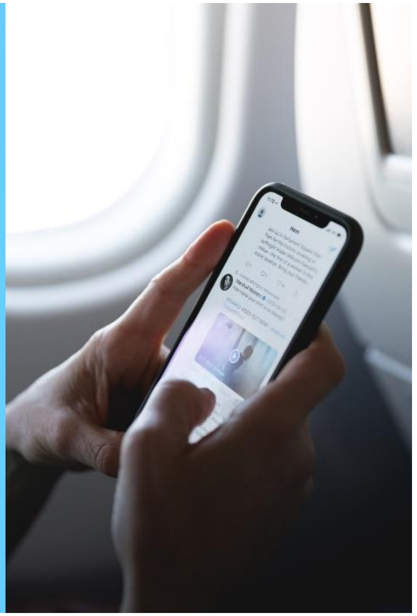
Qual o tamanho amostral necessário para realização do estudo, sabendo que a empresa tem 50 milhões de clientes e um cancelamento de 2% ao mês?



Podemos considerar neste caso, a **população** de clientes da empresa de Telecom como **infinita**.

Adotando uma margem de erro de 0,005 e 95% de confiança, o tamanho amostral é dado por:

$$n = \frac{Z^2 . \hat{p}(1-\hat{p})}{(ME)^2} = \frac{(1{,}96)^2 . 0{,}02\ (0{,}98)}{(0{,}005)^2} = 3.012$$



Realize o cálculo utilizando a fórmula de <u>população finita</u> e discuta as diferenças de resultados.

Uma empresa especializada em cobrança realiza serviços para recuperação de dinheiro para empresas de grande porte. Sabe-se que 7% de todas as tentativas de cobrança resultam no pagamento da dívida.

A empresa utiliza vários canais de contatos: carta, ligação telefônica, SMS, entre outros. Porém, nem todos os canais sempre são utilizados com todos os devedores, devido ao alto custo da operação. O diretor da empresa gostaria de saber, se ele realizasse 100% dos contatos por telefone, seria possível aumentar a % de pagamento, uma vez que este canal é o mais caro e efetivo de todos. Atualmente, o canal telefone é utilizado somente para os clientes com maior chance de pagar.

Para isso, ele gostaria que fosse realizado um teste para uma parcela mínima da base de tal forma que pudesse concluir se realmente este canal alavanca sua % de pagamento.

Qual o tamanho amostral necessário para realização do estudo, sabendo que a empresa precisa recuperar o dinheiro de 2 milhões de devedores, com uma confiança de 95% e margem de erro de 0,01 ?

3.iii. Média: Amostra Estratificada

Uma empresa de aplicativo que fornece serviços de pedreiros, diaristas, faxineiro, passadeira, encanador e etc, gostaria de avaliar a possibilidade de oferecer aos seus prestadores benefícios relacionados a despesas com saúde. A empresa gostaria de saber qual o tamanho amostral para realizar uma pesquisa com seus prestadores, levando em conta que segundo a faixa de idade, os gastos em saúde mudam substancialmente. Considere as seguintes faixas de idade, com seus respectivos N populacionais e S obtidos de uma amostra piloto:

- **Abaixo de 40 anos:** N = 5.000 prestadores, com S=R$120.
- **De 40 a 55 anos:** N = 3.000 prestadores, com S=R$170.
- **Acima de 55 anos:** N = 2.000 prestadores, com S=R$230.

Dado o estrato, todos os clientes possuem a mesma probabilidade de serem sorteados. **Deseja-se estimar o gasto com saúde, considerando uma margem de erro de R$5, com 90% de confiança. Calcule o tamanho amostral total e para cada estrato.**

Definição dos estratos e seus respectivos pesos

Proporção populacional
Soma de W = 1.

Estrato	h	N	%	W
Abaixo de 40 anos	1	5.000	50%	0,50
De 40 a 55 anos	2	3.000	30%	0,30
Acima de 55 anos	3	2.000	20%	0,20
Total		10.000	100%	1

Proporção populacional do Estrato 1 = 0,5.
Proporção populacional do Estrato 2 = 0,3.
Proporção populacional do Estrato 3 = 0,2.

Para estimar o gasto médio com saúde, suponha que retirou-se uma amostra estratificada de **n=2.225** clientes distribuída da seguinte forma:

Estrato	h	N	n
Abaixo de 40 anos	1	5.000	1.113
De 40 a 55 anos	2	3.000	667
Acima de 55 anos	3	2.000	445
Total		10.000	2.225

Veremos mais adiante como obter os tamanhos amostrais por Estrato.

Estimativa pontual populacional

A Tabela apresenta o **peso de cada estrato (W)** e suas respectivas **médias amostrais** do gasto médio com saúde, calculados a partir de **n= 2.225**:

Estrato	**h**	**W**	**X_barra**
Abaixo de 40 anos	**1**	0,5	R$ 250
De 40 a 55 anos	**2**	0,3	R$ 350
Acima de 55 anos	**3**	0,2	R$ 450
Total		1	R$ 320

Qual o valor da **média populacional <u>estimada</u>** pela amostra ?

O **estimador da média populacional**, calculado por meio da amostragem estratificada, é dado por:

$$\bar{X}_{st} = W_1\bar{X}_1 + W_2\bar{X}_2 + W_3\bar{X}_3$$

Em que,

- W_h é o **peso** (proporção populacional) amostral associado ao estrato h.
- $\bar{X}_h$ é a media amostral do estrato h, com h = 1, 2, e 3.

Estimativa pontual populacional

3.iii. MÉDIA: AMOSTRA ESTRATIFICADA | CÁLCULO DA ESTIMATIVA INTERVALAR PARA AMOSTRA ESTRATIFICADA

A Tabela apresenta o **peso de cada estrato (W)** e suas respectivas **médias amostrais** do gasto médio com saúde:

Estrato	h	W	X_barra
Abaixo de 40 anos	**1**	0,5	R$ 250
De 40 a 55 anos	**2**	0,3	R$ 350
Acima de 55 anos	**3**	0,2	R$ 450
Total		1	R$ 320

A **média populacional <u>estimada</u>** é dado por:

$$\bar{X}_{st} = 0{,}5 * 250 + 0{,}3 * 350 + 0{,}2 * 450$$

$$\bar{X}_{st} = R\$320$$

Os prestadores da empresa de aplicativo gastam em média R$320/mês com saúde.

Desvio padrão da média

Existe uma incerteza atrelada à estimativa pontual $\bar{X}_{st}$= R$ 320 que depende do seu desvio padrão.

Qual o <u>desvio padrão de</u> $\bar{X}_{st}$ referente ao gasto em saúde dos prestadores?

Dado que, para **cada estrato** é fornecido o <u>desvio padrão estimado</u> do gasto com saúde, como calcular o **<u>desvio padrão para média amostral</u>** de todos os prestadores?

Estrato	h	W	X_barra	S_estimado
Abaixo de 40 anos	1	0,5	R$ 250	R$ 128
De 40 a 55 anos	2	0,3	R$ 350	R$ 183
Acima de 55 anos	3	0,2	R$ 450	R$ 242
Total		1	R$ 320	

Estimados pela amostra n.

Desvio padrão da média

O desvio padrão é obtido por meio da raiz quadrada da variância.

A variância de $\bar{X}_{st}$ é dada por:

$$V(\bar{X}_{st}) = (W_1)^2 \left(\frac{S_1{}^2}{n_1}\right) . \frac{(N_1 - n_1)}{N_1 - 1} + (W_2)^2 \left(\frac{S_2{}^2}{n_2}\right) . \frac{(N_2 - n_2)}{N_2 - 1} + (W_3)^2 \left(\frac{S_3{}^2}{n_3}\right) . \frac{(N_3 - n_3)}{N_3 - 1}$$

Em que :

- S_1^2 , S_2^2 , S_3^2 são as variâncias amostrais dos estratos.
- **N** é o tamanho da **população**, com **N_h** sendo o tamanho da amostra h, com h = 1, 2 e 3.
- **n** é o tamanho da **amostra**, com n_h sendo o tamanho da amostra h, com h = 1, 2 e 3.
- **W_h** é o **peso** associado ao estrato, com h = 1, 2 e 3.

Variância da média

3.iii. MÉDIA: AMOSTRA ESTRATIFICADA | CÁLCULO DO DESVIO PADRÃO E VARIÂNCIA

A variância de $\bar{X}_{st}$ é dada por:

$$V(\bar{X}_{st}) = (W_1)^2\left(\frac{{S_1}^2}{n_1}\right).\frac{(N_1-n_1)}{N_1-1} + (W_2)^2\left(\frac{{S_2}^2}{n_2}\right).\frac{(N_2-n_2)}{N_2-1} + (W_3)^2\left(\frac{{S_3}^2}{n_3}\right).\frac{(N_3-n_3)}{N_3-1}$$

$$V(\bar{X}_1) = (W_1)^2\left(\frac{{S_1}^2}{n_1}\right).\frac{(N_1-n_1)}{N_1-1} = (0{,}5)^2\left(\frac{128^2}{1.113}\right).\frac{(5.000-1.113)}{5.000-1} = 2{,}86$$

$$V(\bar{X}_2) = (W_2)^2\left(\frac{{S_2}^2}{n_2}\right).\frac{(N_2-n_2)}{N_2-1} = (0{,}3)^2\left(\frac{183^2}{667}\right).\frac{(3.000-667)}{3.000-1} = 3{,}51$$

$$V(\bar{X}_3) = (W_3)^2\left(\frac{{S_3}^2}{n_3}\right).\frac{(N_3-n_3)}{N_3-1} = (0{,}2)^2\left(\frac{242^2}{445}\right).\frac{(2.000-445)}{2.000-1} = 4{,}09$$

h	W	N	n	X_barra		S_estimado	
1	0,5	5.000	1.113	R$	250	R$	128
2	0,3	3.000	667	R$	350	R$	183
3	0,2	2.000	445	R$	450	R$	242
	1	10.000	2.225	R$	320		

A variância de $\bar{X}_{st}$ é dada por:

$$V(\bar{X}_{st}) = (W_1)^2 \left(\frac{S_1{}^2}{n_1}\right) . \frac{(N_1 - n_1)}{N_1 - 1} + (W_2)^2 \left(\frac{S_2{}^2}{n_2}\right) . \frac{(N_2 - n_2)}{N_2 - 1} + (W_3)^2 \left(\frac{S_3{}^2}{n_3}\right) . \frac{(N_3 - n_3)}{N_3 - 1}$$

$$V(\bar{X}_{st}) = \quad 2{,}86 \quad + \quad 3{,}51 \quad + \quad 4{,}09$$

$$V(\bar{X}_{st}) = 10{,}47$$

Estrato	h	W	N	n	X_barra		S_estimado		Variância de X_barra	
Abaixo de 40 anos	1	0,5	5.000	1.113	R$	250	R$	128	R$	2,86
De 40 a 55 anos	2	0,3	3.000	667	R$	350	R$	183	R$	3,51
Acima de 55 anos	3	0,2	2.000	445	R$	450	R$	242	R$	4,09
Total		1	10.000	2.225	R$	320			**R$**	**10,47**

Desvio Padrão $(\bar{X}_{st}) = \text{raiz}(10{,}47)$ = 3,235

O intervalo de confiança para a média populacional, calculado por meio da amostragem estratificada, é dado por:

Intervalo de confiança = $\bar{X}_{st} \pm Z \cdot \sqrt{Var(\bar{X}_{st})}$.

IC = 320 $\pm$ 1,645 $\cdot$ 3,235 = 320 $\pm$5,322

O intervalo com 90% de confiança com gastos em saúde é dado por:

Intervalo de confiança = $(R\$\ 314{,}7; R\$325{,}3)$.

Com 90% de confiança, o verdadeiro gasto médio em saúde dos prestadores está entre R$314,7 e R$325,3.

Case: Estimação do gasto com cartão de crédito

3.iii. MÉDIA: AMOSTRA ESTRATIFICADA | ESTIMADOR PONTUAL

Considere uma Instituição Financeira que deseja estimar o gasto médio com cartão de crédito das famílias de Classes A, B e C de uma determinada agência. Considere que na base de dados tem-se 1.000 clientes da classe A, 2.000 clientes da classe B e de 3.000 clientes da classe C.

Na amostra obteve-se um gasto médio de R$ 12.000,00 para os clientes da classe A, R$ 4.000,00 para os clientes da classe B e de R$ 850,00 para os clientes da classe C.

(a) Obtenha um valor estimado para o gasto médio dos clientes (X_barra), usando a equação:

$$\bar{X}_{st} = W_1\bar{X}_1 + W_2\bar{X}_2 + W_3\bar{X}_3$$

Case: Estimação do gasto com cartão de crédito

3.iii. MÉDIA: AMOSTRA ESTRATIFICADA | DESVIO PADRÃO

Considere uma Instituição Financeira que deseja estimar o gasto médio com cartão de crédito das famílias de Classes A, B e C de uma determinada agência. Considere que na base de dados tem-se 1.000 clientes da classe A, 2.000 clientes da classe B e de 3.000 clientes da classe C.

Na amostra obteve-se os valores da média (X_barra) e desvio padrão (S), com seus respectivos tamanhos amostrais, respectivamente.

Estrato	N	%	W	X_barra	n	S
1 – Classe A	1.000	17%	0,17	R$ 12.000	96	R$ 200
2 – Classe B	2.000	33%	0,33	R$ 4.000	198	R$ 50
3 – Classe C	3.000	50%	0,50	R$ 850	65	R$ 65

(b) Obtenha o valor estimado para o desvio padrão do gasto médio de todas as famílias, utilizando:

$$V(\bar{x}_{st}) = (W_1)^2\left(\frac{{S_1}^2}{n_1}\right).\frac{(N_1 - n_1)}{N_1 - 1} + (W_2)^2\left(\frac{{S_2}^2}{n_2}\right).\frac{(N_2 - n_2)}{N_2 - 1} + (W_3)^2\left(\frac{{S_3}^2}{n_3}\right).\frac{(N_3 - n_3)}{N_3 - 1}$$

Case: Estimação do gasto com cartão de crédito

Considere uma Instituição Financeira que deseja estimar o gasto médio com cartão de crédito das famílias de Classes A, B e C de uma determinada agência. Considere que na base de dados tem-se 1.000 clientes da classe A, 2.000 clientes da classe B e de 3.000 clientes da classe C.

(c) Obtenha um intervalo com 95 % de confiança para o gasto médio, de acordo com a variância encontrada do item (b).

Intervalo de confiança = $\bar{X}_{st} \pm Z \cdot \sqrt{Var(\bar{X}_{st})}$.

Estimativa intervalar: População Infinita

3.iii. MÉDIA: AMOSTRA ESTRATIFICADA | ESTIMATIVA INTERVALAR

O intervalo de confiança para a média populacional, calculado por meio da amostragem estratificada, é dado por:

Intervalo de confiança = $\bar{X}_{st} \pm Z \cdot \sqrt{Var(\bar{X}_{st})}$.

A variância de $\bar{X}_{st}$ no caso de **população infinita ("muito grande")** é dada por:

$$V(\bar{x}_{st}) = \frac{1}{N^2}\left(N_1.\left(\frac{{S_1}^2}{n_1}\right).(N_1 - n_1) + N_2.\left(\frac{{S_2}^2}{n_2}\right).(N_2 - n_2) + N_3.\left(\frac{{S_3}^2}{n_3}\right).(N_3 - n_3)\right)$$

Em que :

- S_1^2 , S_2^2 , S_3^2 são as variâncias amostrais dos estratos.
- **N** é o tamanho da **população**, com **N_h** sendo o tamanho da amostra h, com h = 1, 2 e 3.
- **n** é o tamanho da **amostra**, com $\boldsymbol{n_h}$ sendo o tamanho da amostra h, com h = 1, 2 e 3.

Considere uma Instituição Financeira que deseja estimar o gasto médio com cartão de crédito dos clientes das Classes A, B e C de uma determinada cidade. Considere que na cidade tem-se 200.000 clientes da classe A, 475.000 clientes da classe B e de 885.000 clientes da classe C.

Na amostra obteve-se um gasto médio de R$ 8.200,00 para os clientes da classe A, R$ 3.520,00 para os clientes da classe B e de R$ 850,00 para os clientes da classe C.

(a) Obtenha um valor estimado para o gasto médio **(X_barra)** dos clientes, usando a equação:

$$\bar{X}_{st} = W_1\bar{X}_1 + W_2\bar{X}_2 + W_3\bar{X}_3$$

Case: Estimação do gasto com cartão de crédito

3.iii. MÉDIA: AMOSTRA ESTRATIFICADA | DESVIO PADRÃO

Considere uma Instituição Financeira que deseja estimar o gasto médio com cartão de crédito dos clientes das Classes A, B e C de uma determinada cidade. Considere que na cidade tem-se 200.000 clientes da classe A, 475.000 clientes da classe B e de 885.000 clientes da classe C.

Na amostra obteve-se os valores da média (X_barra) e desvio padrão (S), com seus respectivos tamanhos amostrais, respectivamente.

Estrato	N	X_barra	n	S
1 – Classe A	200.000	R$ 8.200	256	R$ 1.500
2 – Classe B	475.000	R$ 3.520	608	R$ 500
3 – Classe C	885.000	R$ 850	1.134	R$ 200

(b) Obtenha o valor estimado para o desvio padrão do gasto médio de todas as famílias, utilizando:

$$V(\bar{x}_{st}) = \frac{1}{N^2}\left(N_1.\left(\frac{{S_1}^2}{n_1}\right).(N_1 - n_1) + N_2.\left(\frac{{S_2}^2}{n_2}\right).(N_2 - n_2) + N_3.\left(\frac{{S_3}^2}{n_3}\right).(N_3 - n_3)\right)$$

Case: Estimação do gasto com cartão de crédito

Considere uma Instituição Financeira que deseja estimar o gasto médio com cartão de crédito dos clientes das Classes A, B e C de uma determinada cidade. Considere que na cidade tem-se 200.000 clientes da classe A, 475.000 clientes da classe B e de 885.000 clientes da classe C.

(c) Obtenha um intervalo com 95 % de confiança para o gasto médio, de acordo com a variância encontrada do item (b).

Intervalo de confiança = $\bar{X}_{st} \pm Z \cdot \sqrt{Var(\bar{X}_{st})}$.

Ponderação pelo estrato

- Passo 1: Tamanho amostral **geral**, considerando, p.e., 3 estratos:

Alocação Proporcional:

$$n = \frac{N(N_1.S_1^2 + N_2.S_2^2 + N_3.S_3^2)}{N^2.\left(\frac{ME}{Z}\right)^2 + (N_1.S_1^2 + N_2.S_2^2 + N_3.S_3^2)}$$

Alocação de *Neyman*:

$$n = \frac{(N_1.S_1 + N_2.S_2 + N_3.S_3)^2}{N^2.\left(\frac{ME}{Z}\right)^2 + (N_1.S_1^2 + N_2.S_2^2 + N_3.S_3^2)}$$

- Passo 2: Tamanho amostral de **cada estrato** h:

Alocação Proporcional:

$$n_h = \frac{N_h}{N}.n$$

Alocação de *Neyman*:

$$n_h = \frac{N_h S_h}{\sum N_h S_h}.n$$

Maior peso para as maiores dispersões.

Em que :

- **Z** é o valor da **Normal Padrão** que fornece a confiança desejada.
- S_1^2 , S_2^2 , S_3^2 são as **variâncias amostrais** dos estratos a serem estimadas por uma amostra piloto.
- **N** é o tamanho da **população**, com **N_h** sendo o tamanho da amostra h, com h = 1, 2 e 3.
- **n** é o tamanho da **amostra**, com $\boldsymbol{n_h}$ sendo o tamanho da amostra h, com h = 1, 2 e 3.
- **ME** é a **margem de erro**.

Case: Estimação de gastos com saúde

Uma empresa de aplicativo que fornece serviços de pedreiros, diaristas, faxineiro, passadeira, encanador e etc, gostaria de avaliar a possibilidade de oferecer aos seus prestadores benefícios relacionados a despesas com saúde. A empresa gostaria de saber qual o tamanho amostral para realizar uma pesquisa com seus prestadores, levando em conta que segundo a faixa de idade, os gastos em saúde mudam substancialmente. Considere as seguintes faixas de idade, com seus respectivos N populacionais e S obtidos de uma amostra piloto:

- **Abaixo de 40 anos:** N = 5.000 prestadores, com S=R$120.
- **De 40 a 55 anos:** N = 3.000 prestadores, com S=R$170.
- **Acima de 55 anos:** N = 2.000 prestadores, com S=R$230.

Dado o estrato, todos os clientes possuem a mesma probabilidade de serem sorteados. **Deseja-se estimar o gasto com saúde, considerando uma margem de erro de R$5, com 90% de confiança. Calcule o tamanho amostral total e para cada estrato.**

Case: Estimação de gastos com saúde

3.iii. MÉDIA: AMOSTRA ESTRATIFICADA | ALOCAÇÃO PROPORCIONAL

O tamanho amostral para a amostra estratificada com 90% de confiança utilizando **alocação proporcional** é dado por:

$$n = \frac{N(N_1.S_1^2 + N_2.S_2^2 + N_3.S_3^2)}{N^2.\left(\frac{ME}{Z}\right)^2 + (N_1.S_1^2 + N_2.S_2^2 + N_3.S_3^2)}$$

$$n = \frac{(10.000)(5.000\,.\,120^2 \; + 3.000\,.170^2 + 2.000\,.\,230^2)}{(10.000)^2\left(\frac{5}{1{,}645}\right)^2 + (5.000.\,120^2 + 3.000.\,170^2 + 2.000.\,230^2)}$$

$$n = 2.225$$

A quantidade de prestadores necessários para pesquisa de gastos com a saúde, com 90% de confiança e margem de erro R$5, seria **2.225**, correspondente a aproximadamente **22% da população** de prestadores da empresa.

O tamanho amostral pela **alocação proporcional** seria:

$$n_h = \frac{N_h}{N} . n$$

$$n_1 = \frac{5.000}{10.000} . n = 0{,}5 \;.\; 2.225 = 1.113$$

n_1 corresponde a 50% da amostra n.

$$n_2 = \frac{3.000}{10.000} . n = 0{,}3 \;.\; 2.225 = 667$$

n_2 corresponde a 30% da amostra n.

$$n_3 = \frac{2.000}{10.000} . n = 0{,}2 \;.\; 2.225 = 445$$

n_3 corresponde a 20% da amostra n.

O tamanho amostral para a amostra estratificada com 90% de confiança utilizando **alocação de Neyman** é dado por:

$$n = \frac{(N_1.S_1 + N_2.S_2 + N_3.S_3)^2}{N^2.\left(\frac{ME}{Z}\right)^2 + (N_1.S_1^2 + N_2.S_2^2 + N_3.S_3^2)}$$

$$n = \frac{(5.000\,.\,120\; + 3.000\,.170 + 2.000\,.\,230)^2}{(10.000)^2\left(\frac{5}{1{,}645}\right)^2 + (5.000.\,120^2 + 3.000.\,170^2 + 2.000.\,230^2)}$$

$$n = 2.074$$

A quantidade de prestadores necessários para pesquisa de gastos com a saúde, com 90% de confiança e margem de erro R$5, seria **2.074**, correspondente a quase **21% da população** de prestadores da empresa.

O tamanho amostral pela **alocação de *Neyman*** seria:

$$n_h = \frac{N_h S_h}{\sum N_h S_h} . n$$

$$n_1 = \frac{5.000 \,.\, 120}{5.000 \,.\, 120 + 3.000 \,.\, 170 + 2.000 \,.\, 230} . 2.074 = 0{,}38 \,.\, 2.074 = 792$$

n_1 corresponde a 38% da amostra n.

$$n_2 = \frac{3.000 \,.\, 170}{5.000 \,.\, 120 + 3.000 \,.\, 170 + 2.000 \,.\, 230} . 2.074 = 0{,}32 \,.\, 2.074 = 674$$

n_2 corresponde a 32% da amostra n.

$$n_3 = \frac{2.000 \,.\, 230}{5.000 \,.\, 120 + 3.000 \,.\, 170 + 2.000 \,.\, 230} . 2.074 = 0{,}29 \,.\, 2.074 = 608$$

n_3 corresponde a 29% da amostra n.

Case: Estimação de gastos com saúde

3.iii. MÉDIA: AMOSTRA ESTRATIFICADA | CÁLCULO DO TAMANHO AMOSTRAL

Considerando uma margem de erro de R$5, 90% de confiança, para uma população de 10.000 prestadores, obtém-se o **tamanho amostral total** e **por estrato**:

Alocação Proporcional

Alocação de Neyman

Estrato	N	%	W	S_piloto	n_proporcional	%_proporcional	n_Neyman	%_Neyman
1	5.000	50%	0,50	R$ 120	1.113	50%	792	38%
2	3.000	30%	0,30	R$ 170	667	30%	674	32%
3	2.000	20%	0,20	R$ 230	445	20%	608	29%
Total	10.000	100%	1		2.225		2.074	

- Note que a alocação proporcional é equivalente a uma AAS no caso quando se tem acesso a 100% da população.
- A alocação de *Neyman* utiliza uma ponderação em que requer mais elementos quando há uma maior variabilidade nos estratos.

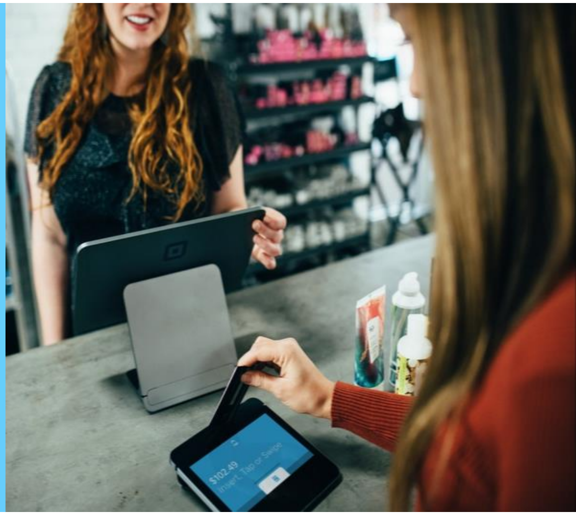
Case: Estimação de gasto mensal

Uma grande rede de varejos deseja estimar o gasto mensal dos seus clientes por classe social. Para isso gostaria de realizar uma pesquisa com uma pequena parte de seus clientes para realizar tal estimativa.

Classe	N populacional	S Piloto
A	1.000	R$ 100
B	2.000	R$ 70
C	3.000	R$ 50

Pesquisa →

Média Amostral	S Amostral
R$ 5.000	R$ 120
R$ 1.700	R$ 75
R$ 700	R$ 55



(a) Calcule o tamanho amostral considerando uma confiança de 95% e margem de erro R$5.
(b) Calcule o n amostral por estrato considerando **alocação proporcional e *Neyman***.
(c) Após a pesquisa realizada, obteve-se os resultados da "Pesquisa" e foram calculados as médias e DP (desvio padrão) amostrais por estrato. Calcule a estimativa pontual da média do gasto mensal dos clientes da rede varejista.
(d) Calcule o intervalo com 95% confiança para o gasto médio dos clientes da rede varejista, considerando:

$$V(\bar{x}_{st}) = (W_1)^2\left(\frac{{S_1}^2}{n_1}\right).\frac{(N_1-n_1)}{N_1-1} + (W_2)^2\left(\frac{{S_2}^2}{n_2}\right).\frac{(N_2-n_2)}{N_2-1} + (W_3)^2\left(\frac{{S_3}^2}{n_3}\right).\frac{(N_3-n_3)}{N_3-1}$$

4. Exercícios

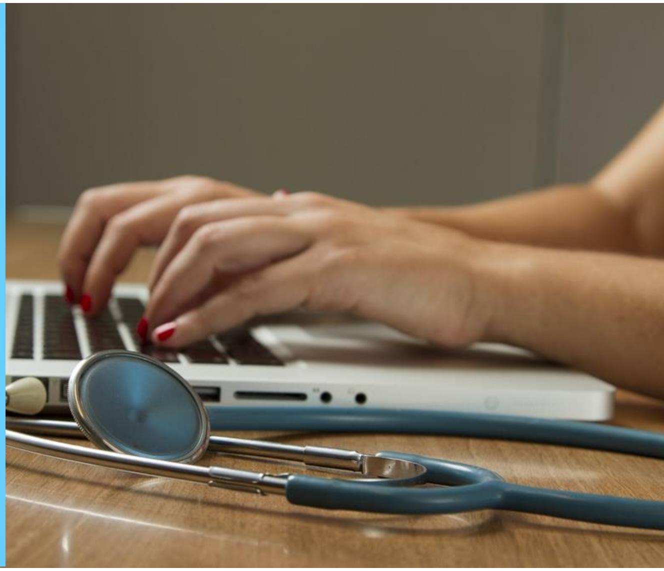
Case: Atendimento de corretores

Uma construtora deseja estimar o tempo médio de atendimento de um cliente pelo corretor de imóveis. Para isso, gostaria de conduzir uma pesquisa com uma pequena amostra de corretores. Obtenha a **quantidade de corretores necessários para pesquisa**, considerando uma margem de erro de 3 minutos, confiança de 95% e desvio padrão (S) de 20 minutos. A quantidade total de corretores da construtora é N=500.

Case: Atendimento em clínica médica

Uma grande clínica médica que atende aproximadamente 1.000 pacientes por mês, e estima que 30% dos pacientes não voltarão mais para uma próxima consulta. Desta forma, gostaria de confirmar esta suspeita, realizar uma pesquisa para estimar o percentual de “não retornos” e traçar estratégias para que o paciente retorne para novas consultas.

Qual a quantidade de pacientes que a clínica precisaria entrevistar, adotando uma margem de erro de 3 p.p. e confiança de 95%?



Case: Estimação valor gasto em cesta básica

Considere que o gestor de um supermercado deseja calcular o **tamanho da amostra** para estimar o **gasto com cesta básica** de acordo com classe social. Considere para o cálculo do tamanho amostral margem de erro de R$5, 95% de confiança, S e N dados a seguir:

Estrato	N	W	S
1 - Classe A	200	0,05	R$ 300
2 - Classe B	500	0,14	R$ 200
3 - Classe C	3.000	0,81	R$ 100
Total	3.700		

(a) Calcule o tamanho amostral por estrato considerando **alocação proporcional**.
(b) Calcule o tamanho amostral por estrato considerando **alocação de *Neyman***.

Referências Bibliográficas

1. Anderson, R. A., Sweeney, J. D. e Williams, T. A. *Estatística Aplicada à Administração e Economia.* Editora Cengage. 4ª edição, 2019.

2. Rao, P. S. *Sampling methodologies with applications*. Chapman and Hall/CRC, 2000.

3. Kish, L. *Survey sampling*. John Wiley & Sons, Inc., 6th Edition, 1965.