PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE TRANSPORTES E GESTÃO TERRITORIAL – PPGTG
DEPARTAMENTO DE ENGENHARIA CIVIL – ECV

DISCIPLINA: TGT410026 – FUNDAMENTOS DE ESTATÍSTICA

7ª AULA: AMOSTRAGEM E DISTRIBUIÇÕES AMOSTRAIS

Iniciamos a aula com a reapresentação das definições de:

População – conjunto de todos os elementos de interesse em um estudo.

Amostra – é um subconjunto da população.

Características numéricas da população: média (μ) e desvio padrão (σ), além da proporção (p).

Características numéricas da amostra: média ($\bar{x}$) e desvio padrão (s), além da proporção amostral ($\bar{p}$).

Importante é o entendimento de que os resultados da amostra fornecem ***apenas estimativas*** dos valores das características da população.

E qual a ***confiança*** que as estimativas resultem em "bons" resultados? A técnica de amostragem e o conhecimento da distribuição amostral apropriada nos permitem responder a esta questão.

AMOSTRAGEM ALEATÓRIA SIMPLES

A definição de amostra aleatória simples e o processo de seleção de uma amostra aleatória simples dependem de a população ser *finita* ou *infinita*.

1. POPULAÇÃO FINITA – uma amostra aleatória simples de tamanho n de uma *população finita* de tamanho N é uma amostra selecionada de tal maneira que cada amostra possível de tamanho n tenha a mesma probabilidade de ser escolhida.

Em geral, para selecionarmos amostras aleatórias simples de uma população finita, fazemos uso da tabela de números aleatórios. Verificar o uso dos números aleatórios em um exemplo.

Tabela de nºs aleatórios (com 5 dígitos)
63271 59986 71744 51102 15141 80714 58683 93108...
Da esquerda para a direita os sete primeiros números aleatórios de 4 dígitos são:
6327 1599 8671 7445 1102 1514 1807
Da esquerda para a direita os cinco primeiros números aleatórios de 2 dígitos são:
63 27 15 99 86
e assim por diante.

2. POPULAÇÃO INFINITA - existem casos em que a população é infinita ou tão grande que, para fins práticos, precisa ser tratada como infinita. Como exemplo, suponha um restaurante de *fast-food*; o processo contínuo de visitas de clientes ao restaurante pode ser visto como proveniente de uma população infinita.

Uma amostra aleatória simples de uma população infinita é uma amostra selecionada de tal maneira que as condições a seguir sejam satisfeitas:

- cada elemento selecionado vem dessa população;
- cada elemento é selecionado de maneira independente.

CUIDADO!!!!
Retornando ao exemplo do restaurante de *fast-food*. Qualquer cliente que entre no restaurante satisfaz a primeira condição. O objetivo da segunda condição é evitar que ocorra um viés na seleção. Ocorreria um viés de seleção se, por exemplo, cinco clientes consecutivos selecionados fossem, todos, amigos entre si que chegassem junto ao restaurante. Poder-se-ia esperar que esses clientes apresentassem perfis semelhantes. O viés deve ser evitado assegurando-se que cada cliente seja selecionado de forma independente, ou seja, que a escolha de um cliente em particular não influa na escolha de outro cliente qualquer.
O McDonald´s, líder mundial no ramo de restaurantes *fast-food* implementou um sistema de amostragem aleatória simples exatamente para esse tipo de situação: alguns clientes do McDonald´s apresentavam cupons de desconto. Toda vez que a empresa quisesse que o cliente apresentasse cupon de descontos, o cliente era servido e, em seguida, solicitado a preencher um questionário de perfil do cliente.
Populações infinitas frequentemente estão associadas a processos ininterruptos que operam continuamente ao longo do tempo. Por

exemplo: peças que são manufaturadas em uma linha de produção; transações financeiras que ocorrem em um banco; as chamadas telefônicas que chegam a um centro de suporte técnico; clientes que entram numa loja; etc.

NOTA

O número de diferentes amostras aleatórias simples de tamanho n que podem ser selecionadas de uma população infinita de tamanho N é

$$C_n^N = \frac{N!}{n!(N-n)!}$$

EXERCÍCIOS

1. Considere que uma população finita tenha 350 elementos. Usando os três últimos dígitos de cada um dos seguintes números aleatórios de cinco dígitos apresentados a seguir, determine os quatro primeiros elementos que serão selecionados para a amostra aleatória simples.

Nºs aleatórios de cinco dígitos:

98601 73022 83448 02147 34229 27553 84147 93289 14209

Resposta:

O tamanho da amostra é: $n = 4$; usar os três últimos dígitos dos nºs aleatórios acima para selecionar os quatro elementos que farão parte da amostra. Observe que o tamanho da população é $N = 350$; portanto, somente nºs aleatórios $\leq$ 350 serão considerados. Assim, os nºs selecionados são: 022 147 229 289.

2. Os dez títulos financeiros mais ativos nas Bolsas de New York (Nyse), Nasdaq e American (Amex) com capitalização de mercado acima de US$ 500 milhões são os seguintes:

1) Applied Material	2) Cisco Systems	3) Intel
4) Lucent Technologies	5) Microsoft	6) Nasdaq 100
7) Nextel	8) Oracle	9) SPDR
10) Sun Microsystems		

a) suponha que uma amostra aleatória de cinco títulos financeiros seja selecionada para um estudo detalhado do comportamento dos negócios. Iniciando com o primeiro dígito aleatório da tabela de nºs aleatórios com 5 dígitos e lendo a coluna de cima para baixo, use os nºs aleatórios de um único dígito aleatório para selecionar uma amostra aleatória simples de 5 títulos financeiros a serem usados nesse estudo;

b) de acordo com a Nota acima, quantas amostras aleatórias simples de tamanho 5 podem ser selecionadas da lista de dez títulos

financeiros?
Resposta:
a) Nasdaq; Oracle; Microsoft; Lucent; Applied
b) Devemos calcular a combinação de N = 10 e n = 5.

$$C_5^{10} = \frac{10!}{5!(10-5)!} = \frac{10!}{5!.5!} = 252 \text{ amostras.}$$

3. Indique se as populações a seguir devem ser consideradas *finitas* ou *infinitas*:
a) Todos os eleitores inscritos do estado de Santa Catarina.
b) Todos os aparelhos de televisão que poderiam ser produzidos pelo parque industrial de Manaus.
c) Todos os pedidos que poderiam ser processados por uma empresa de encomenda postal.
d) Todas as chamadas telefônicas de emergências que poderiam ser feitas a uma delegacia de polícia local.
e) Todos os componentes que a Intelbras produziu no segundo turno de trabalho no dia 17 de maio de 2014.

4. A relação a seguir apresenta os 25 melhores times de futebol americano da temporada 2002. Use a nona coluna dos números aleatórios que se inicia com 13554 para selecionar uma amostra aleatória simples de seis times de futebol. Use os dois primeiros dígitos de cada linha da nona coluna para realizar o seu processo de seleção. Quais os times selecionados?

1.Ohio	6. Kansas	11. Carolina	16.Auburn	21. Colorado
2. Miami	7. Texas	12. Boise Ste	17.NDame	22. TCU
3. Georgia	8. Iowa	13. Maryland	18.Pittsburg	23. Florida Ste
4. California	9. Michigan	14. VTech	19.Marshall	24. Florida
5. Oklahoma	10. WSte	15. Penn Ste	20.WVirginia	25. Virginia

5. Considere uma população finita com cinco elementos rotulados A, B, C, D e E. Dez possíveis amostras aleatórias simples de tamanho 2 podem ser selecionadas.
a) relacione as dez amostras, iniciando com AB, AC e assim por diante.
b) usando a amostragem aleatória simples, qual é a probabilidade de cada amostra de tamanho 2 ser selecionada?
c) considere que o número aleatório 1 corresponde a A, o número aleatório 2 corresponde a B e assim por diante. Relacione a amostra aleatória simples de tamanho 2 que será selecionada usando-se os dígitos aleatórios 8 0 5 7 5 3 2.

ESTIMAÇÃO POR PONTO

Para estimar o valor do parâmetro de uma população calculamos uma característica correspondente da amostra, denominada estatística amostral.
Como exemplo, para estimar a média (μ) e o desvio padrão (σ) de uma população calculamos as estatísticas amostrais correspondentes: a média amostral ($\bar{x}$) e o desvio padrão amostral (s). Podemos ainda estimar a proporção (p) da população usando a proporção amostral ($\bar{p}$).

Relembrando, as expressões para o cálculo são:

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

$$s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}}$$

$$\bar{p} = \frac{x}{n}$$

Ao realizar os cálculos utilizando as expressões acima, executamos o procedimento estatístico denominado ***estimação por ponto***.
A média amostral ($\bar{x}$) é a estimação por ponto da média populacional (μ); o desvio padrão amostral (s) é o estimador por ponto do desvio padrão populacional (σ) e, a proporção amostral ($\bar{p}$) é o estimador por ponto da proporção (p) da população.

É de se esperar que a estimação por ponto difira bastante dos parâmetros populacionais correspondentes. Essa diferença (veremos num exercício) deve ser esperada porque é usada uma amostra e não um censo de toda a população para desenvolver a estimação por ponto.

Aplicação:

Uma pergunta de uma pesquisa realizada com uma amostra de 150 indivíduos produziu 75 respostas “Sim”, 55 respostas “Não” e 20 respostas “Sem opinião”.
a) qual é a estimação por ponto da proporção da população que respondeu “Sim”?
b) qual é a estimação por ponto da proporção da população que respondeu “Não”?

a) $\bar{p} = \frac{x}{n} = \frac{75}{150} = 0{,}50$

b) $\bar{p} = \frac{x}{n} = \frac{55}{150} = 0{,}37$

DISTRIBUIÇÕES AMOSTRAIS

Vamos fundamentar nosso estudo a respeito da distribuição amostral, usando um exemplo real.

O diretor de pessoal da *Electronics Associates, Inc. (EAI)* foi incumbido de desenvolver um perfil dos 2500 gerentes da empresa. As características a serem identificadas incluem o salário médio anual dos gerentes e a proporção de gerentes que concluíram o programa de treinamento gerencial da empresa.

Usando os 2500 gerentes como a população para esse estudo, pode-se encontrar o salário médio anual e o *status* do programa de treinamento de cada indivíduo consultando os registros de pessoal da empresa. Usando o conjunto de dados da *EAI* e as fórmulas já conhecidas, a média populacional e o desvio padrão correspondentes aos dados de salário anual foram computados:

Média populacional: $\mu = US\$51.800$
Desvio padrão da população: $\sigma = US\$4.000$

Os dados referentes ao status no programa de treinamento mostram que 1500 gerentes concluíram o programa de treinamento. Assim, a proporção da população de gerentes que concluiu o programa de treinamento é: $\bar{p} = \frac{1500}{2500} = 0,60$.

Assim, μ = US\$ 51.800, σ = US\$ 4.000 e $\bar{p}$ = 0,60 são os parâmetros da população de gerentes da *EAI*.

Suponha agora que todas as informações necessárias sobre todos os gerentes da *EAI* não estivessem prontamente disponíveis no banco de dados da empresa. A questão a considerar agora é como o diretor de pessoal da empresa pode obter estimativas dos parâmetros da população de gerentes usando uma amostra, em vez de todos os 2500 gerentes da população.

Imagine a decisão de coletar e usar os dados referentes a uma amostra de 30 gerentes. Já sabemos como fazer isso! Lembrar que, entre os 2500 gerentes da empresa, $2,74 \times 10^{69}$ amostras aleatórias simples de 30 gerentes da *EAI* podem ser obtidas.

Foi, então, selecionada uma amostra aleatória simples de 30 gerentes e coletados os registros referentes ao salário anual e a situação no programa de treinamento gerencial de cada gerente selecionado. Os dados estão registrados na Tabela a seguir:

Tabela: Salários anuais e situação no programa de treinamento gerencial referente a uma amostra aleatória simples de 30 gerentes da *EAI*

Salário anual (US$)	Programa de Treinamento Gerencial	Salário anual (US$)	Programa de Treinamento Gerencial
x1 = 49.094,30	Sim	x16 = 51.766,00	Sim
x2 = 53.263,90	Sim	x17 = 52.541,30	Não
x3 = 49.643,50	Sim	x18 = 44.980,00	Sim
x4 = 49.894,90	Sim	x19 = 51.932,60	Sim
x5 = 47.621,60	Não	x20 = 52.973,00	Sim
x6 = 55.924,00	Sim	x21 = 45.120,90	Sim
x7 = 49.092,30	Sim	x22 = 51.753,00	Sim
x8 = 51.404,40	Sim	x23 = 54.391,80	Não
x9 = 50.957,70	Sim	x24 = 50.164,20	Não
x10 = 55.109,70	Sim	x25 = 52.973,60	Não
x11 = 45.922,60	Sim	x26 = 50.241,30	Não
x12 = 57.268,40	Não	x27 = 52.793,90	Não
x13 = 55.688,80	Sim	x28 = 50.979,40	Sim
x14 = 51.564,70	Não	x29 = 55.860,90	Sim
x15 = 56.188,20	Não	x30 = 57.309,10	Não

As estatísticas amostrais calculadas com os dados da tabela resultaram nos seguintes valores:

$$\bar{x} = \frac{\sum x_i}{n} = \frac{1.554.420}{30} = 51.814$$

$$s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{325.009.260}{29}} = 3.348$$

$$\bar{p} = \frac{x}{n} = \frac{19}{30} = 0,63$$

Desse modo, em relação à amostra aleatória simples de 30 gerentes da *EAI* apresentada na Tabela, US$ 51.814 é a estimação por ponto de μ; US$ 3.348 é a estimação por ponto do desvio padrão σ, e 0,63 é a estimação por ponto da proporção *p*.

Resumo das estimações por ponto obtidas de uma amostra aleatória simples de 30 gerentes da EAI

Parâmetro Populacional	Estimativa por ponto
μ = salário médio anual – US$ 51.800	$\bar{x}$ = salário médio anual da amostra – US$ 51.814
σ = desvio padrão do salário anual da população – US$ 4.000	*s* = salário médio anual da amostra – US$ 3.348
p = proporção da população que concluiu o PTG – 0,60	$\bar{p}$ = proporção da amostra que concluiu o PTG - 0,63

O resumo mostra claramente que a estimação por ponto difere bastante dos parâmetros populacionais correspondentes. Já sabemos a razão disso!

ESTAMOS PRONTOS para estudar o comportamento dos valores das estatísticas amostrais utilizadas na estimação dos parâmetros populacionais.

Vamos selecionar uma outra amostra aleatória simples de 30 gerentes da EAI; calculamos as estimações por ponto:

Média da amostra: $\bar{x} = 52.670$

Proporção da amostra: $\bar{p} = 0,70$

Observe que os valores produzidos por essa amostra diferem da primeira amostra. De fato, não se pode esperar que uma segunda amostra aleatória simples de 30 gerentes da *EAI* produza as mesmas estimações por ponto que a primeira.

Vamos repetir diversas vezes o processo de selecionar uma amostra aleatória simples de 30 gerentes da *EAI*, calculando a cada vez os valores de $\bar{x}$ e $\bar{p}$. A tabela a seguir contém uma parte dos resultados obtidos para 500 amostras aleatórias simples:

Tabela: valores de $\bar{x}$ e $\bar{p}$ em 500 amostras aleatórias simples de 30 gerentes da *EAI*

Número da amostra	Média da amostra $\bar{x}$	Proporção da amostra $\bar{p}$
1	51.814	0,63
2	52.670	0,70
3	51.780	0,67
4	51.588	0,53
...	...	...
500	51.752	0,50

A tabela a seguir organiza os dados e fornece a distribuição de frequência e frequência relativa dos 500 valores de $\bar{x}$.

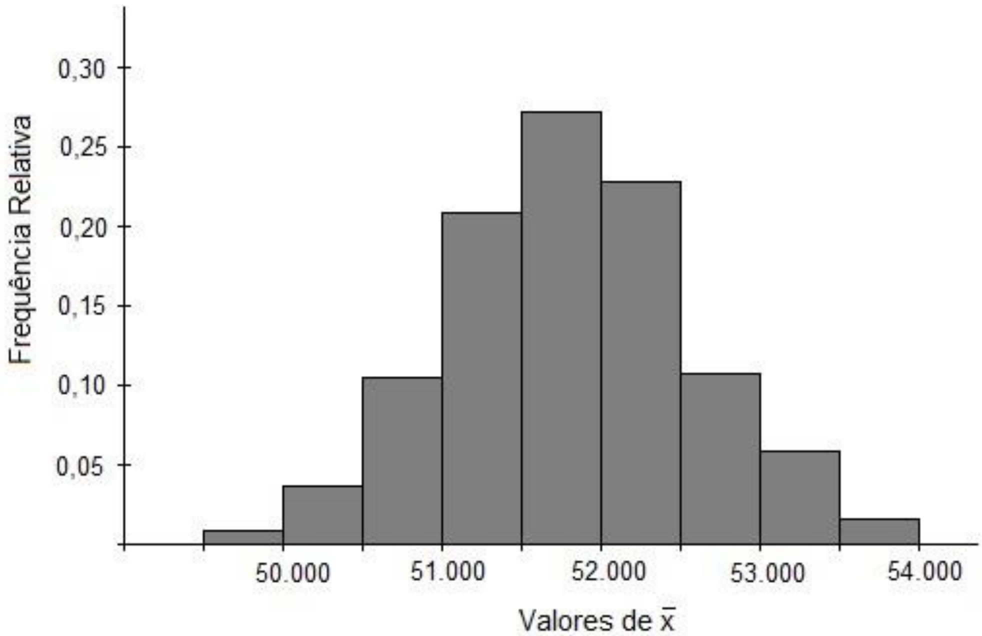
Já definimos uma variável aleatória como uma descrição numérica do resultado de um experimento. SE CONSIDERARMOS QUE O PROCESSO DE SELECIONAR UMA AMOSTRA ALEATÓRIA SIMPLES É UM EXPERIMENTO, a média amostral $\bar{x}$ é uma descrição numérica do resultado do experimento. Desse modo, a média amostral é uma variável aleatória e, a exemplo do que ocorre com qualquer variável aleatória, $\bar{x}$ tem um valor médio ou esperado, um desvio padrão e uma distribuição de probabilidade.

Tabela: Distribuição de frequência de $\bar{x}$ em 500 amostras aleatórias simples de 30 gerentes da *EAI*

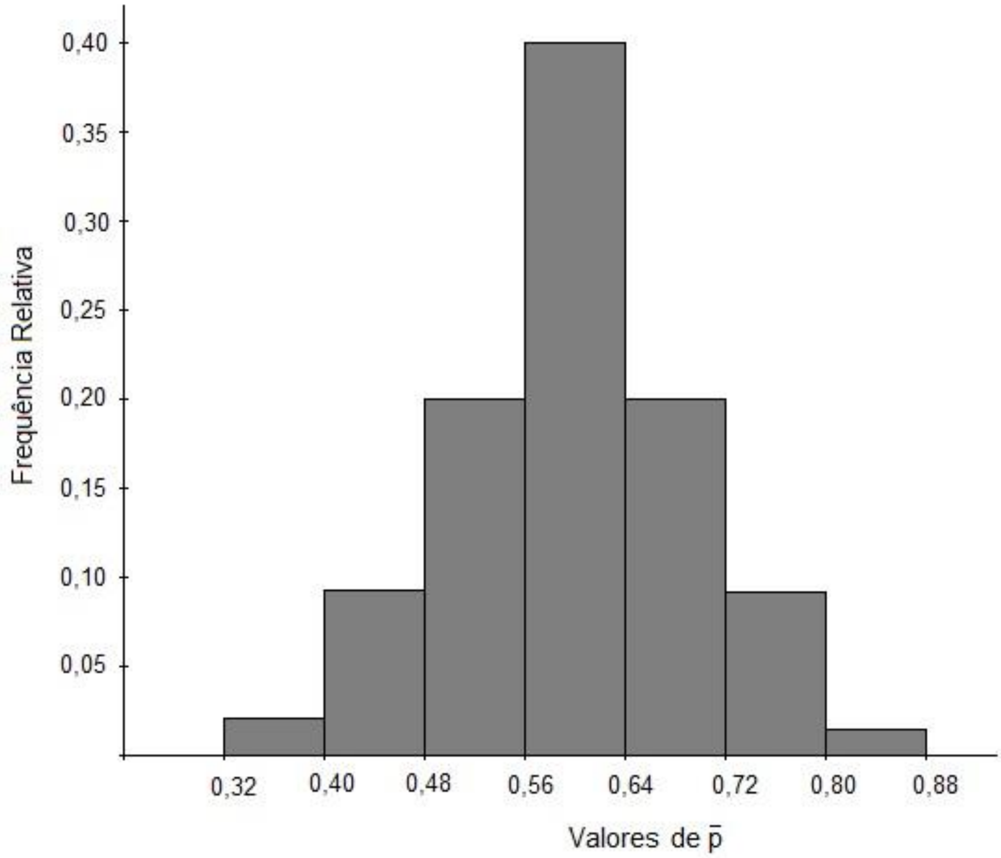
Salário anual médio	Frequência	Frequência relativa
49.500,00 – 49.999,99	2	0,004
50.000,00 – 50.499,99	16	0,032
50.500,00 – 50.999,99	52	0,104
51.000,00 – 51.499,99	101	0,202
51.500,00 – 51.999,99	133	0,266
52.000,00 – 52.499,99	110	0,220
52.500,00 – 52.999,99	54	0,108
53.000,00 – 53.499,99	26	0,052
53.500,00 – 53.999,99	6	0,012
Totais:	500	1,000

Uma vez que os diversos valores possíveis de $\bar{x}$ resultam de diferentes amostras aleatórias simples, a distribuição da probabilidade de $\bar{x}$ é conhecida por ***distribuição amostral*** de $\bar{x}$.
Conhecer essa distribuição amostral e suas características nos possibilitará fazer afirmações a respeito de quão próxima a média da amostra está da média da população.

As figuras a seguir sintetizam os 500 valores de $\bar{x}$ e os 500 valores da proporção da amostra. A forma da distribuição dos valores já nos informa que a distribuição tem a forma de sino e isso nos remete à distribuição normal de probabilidade.



Assim como ocorre com $\bar{x}$, $\bar{p}$ também é uma variável aleatória. O histograma de frequência relativa dos 500 valores da amostra nos dá uma idéia da distribuição amostral de $\bar{p}$.



DISTRIBUIÇÃO AMOSTRAL DE $\bar{x}$

A distribuição amostral de $\bar{x}$ é a distribuição de probabilidade de todos os valores possíveis da média amostral $\bar{x}$. A distribuição amostral de $\bar{x}$ tem um valor esperado (ou média), um desvio padrão e um formato (ou forma) característico.

VALOR ESPERADO DE $\bar{x}$

O valor esperado de $\bar{x}$ é a média de todos os valores possíveis de $\bar{x}$. Assim, $\boldsymbol{E(\bar{x}) = \mu}$ onde μ é a média da população.

DESVIO PADRÃO DE $\bar{x}$

Para definir o desvio padrão da distribuição amostral de $\bar{x}$, precisamos distinguir entre população finita ou infinita. Considerando:

$\sigma_{\bar{x}}$ o desvio padrão da média amostral $\bar{x}$;
σ o desvio padrão da população;
n o tamanho da amostra;

N o tamanho da população;

O desvio padrão de $\bar{x}$ para população finita é calculado por:

$$\sigma_{\bar{x}} = \sqrt{\frac{N-n}{N-1}}.\left(\frac{\sigma}{\sqrt{n}}\right)$$

e para população infinita:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

Verifica-se que no caso de população finita o termo $\sqrt{\frac{N-n}{N-1}}$ é necessário, pois serve para corrigir a relação entre população "grande" e tamanho da amostra relativamente "pequeno". Nesses casos o fator de correção para populações finitas está próximo de 1 e consequentemente a diferença entre $\sigma_{\bar{x}}$ para populações finitas ou infinitas torna-se desprezível.

Diretriz geral para o cálculo do desvio padrão da média amostral:

Use a expressão $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$

Sempre que:
1) a população for infinita;
2) a população for finita e o tamanho da amostra for menor que 5% do tamanho da população; ou seja, $\frac{n}{N} \leq 0,05$.

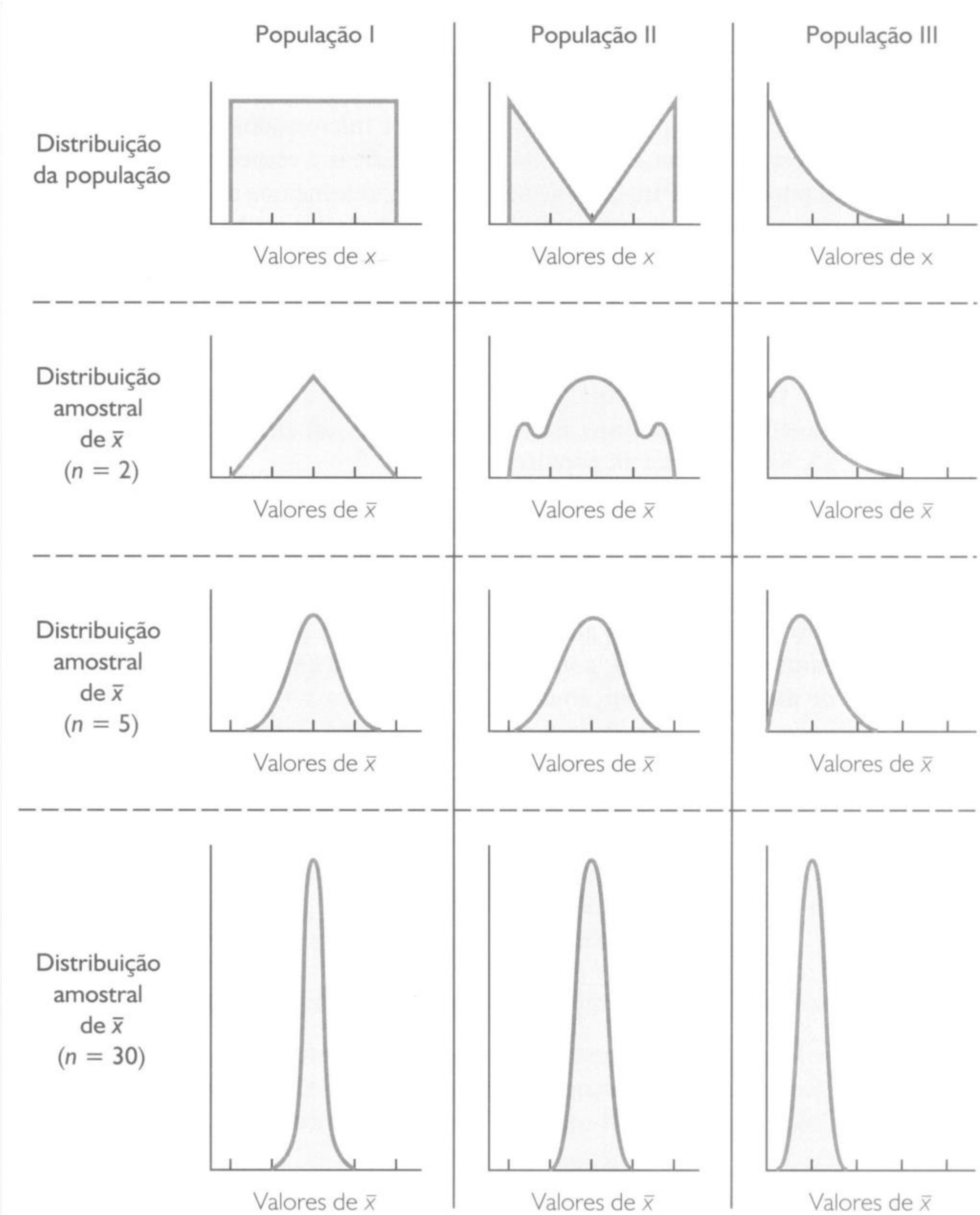
Para o exemplo da *EAI*, temos que N = 2.500 e n = 30. A população é finita e a relação n / N = 0,012. Como o tamanho da amostra é menor que 5% do tamanho da população podemos ignorar o fator de correção para população finita e calcular o desvio padrão da média: $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{4000}{\sqrt{30}} = 730,3$.

O desvio padrão da média amostral $\sigma_{\bar{x}}$ recebe a denominação de ***erro padrão*** da média. O erro padrão da média vai ser útil para determinarmos o quão distante a média amostral pode estar da média populacional.

FORMA DA DISTRIBUIÇÃO AMOSTRAL DE $\bar{x}$

Considerar dois casos:
a) ***A população tem uma distribuição normal*** – quando a população tem distribuição normal, a distribuição amostral de $\bar{x}$ está normalmente distribuída para qualquer tamanho de amostra.

b) ***A população não tem uma distribuição normal*** – nesse caso, devemos apelar para o teorema do limite central que diz: "ao selecionar amostras aleatórias simples de tamanho *n* de uma população, podemos aproximar a distribuição amostral da média da amostra $\bar{x}$ por meio de uma distribuição normal à medida que aumentamos o tamanho da amostra".

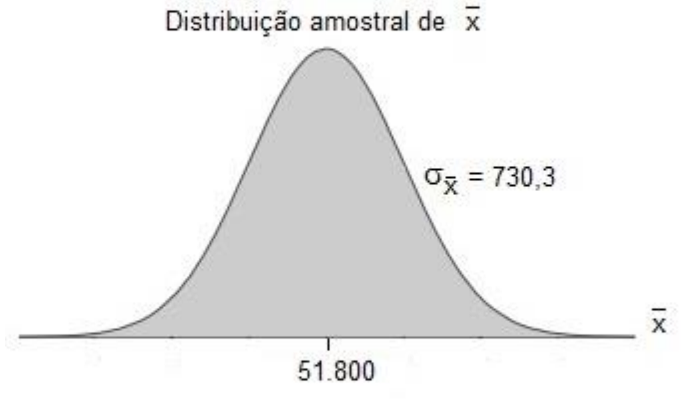


A figura mostra o funcionamento do teorema do limite central em relação a três populações diferentes: a população I segue uma distribuição uniforme; a população II segue uma distribuição denominada "orelha-de-coelho"; a população III tem distribuição similar à distribuição exponencial. Nenhuma delas passa, nem perto, da distribuição normal.

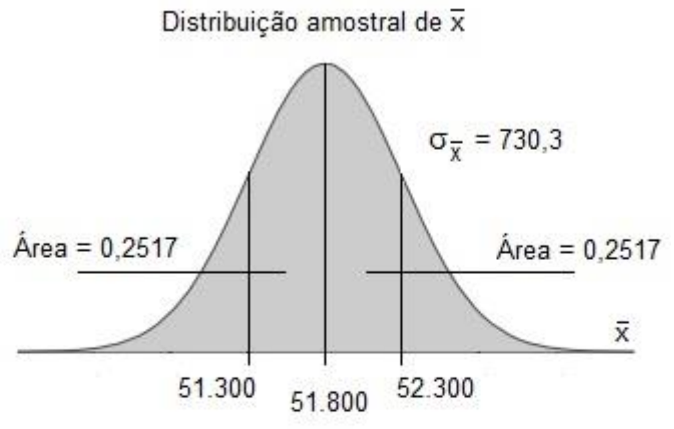
As três linhas seguintes da figura mostram a forma da distribuição amostral à medida que o tamanho da amostra aumenta. Observe que para amostras de tamanho n = 30, as formas das distribuições são aproximadamente normais.

VALOR PRÁTICO DA DISTRIBUIÇÃO AMOSTRAL DE $\bar{x}$

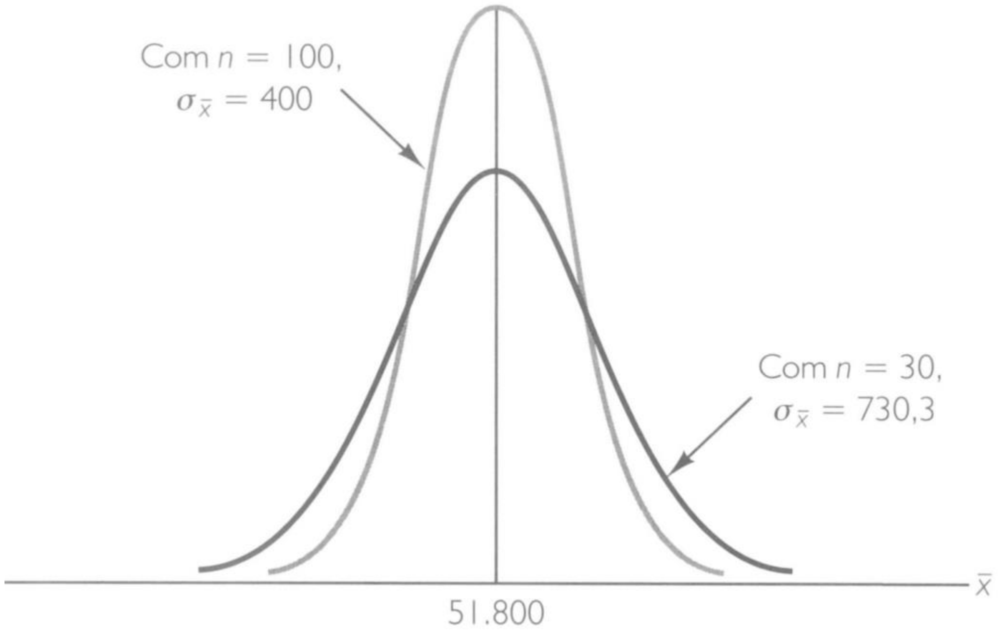
Distribuição amostral do salário médio anual de uma amostra aleatória simples de 30 gerentes da EAI

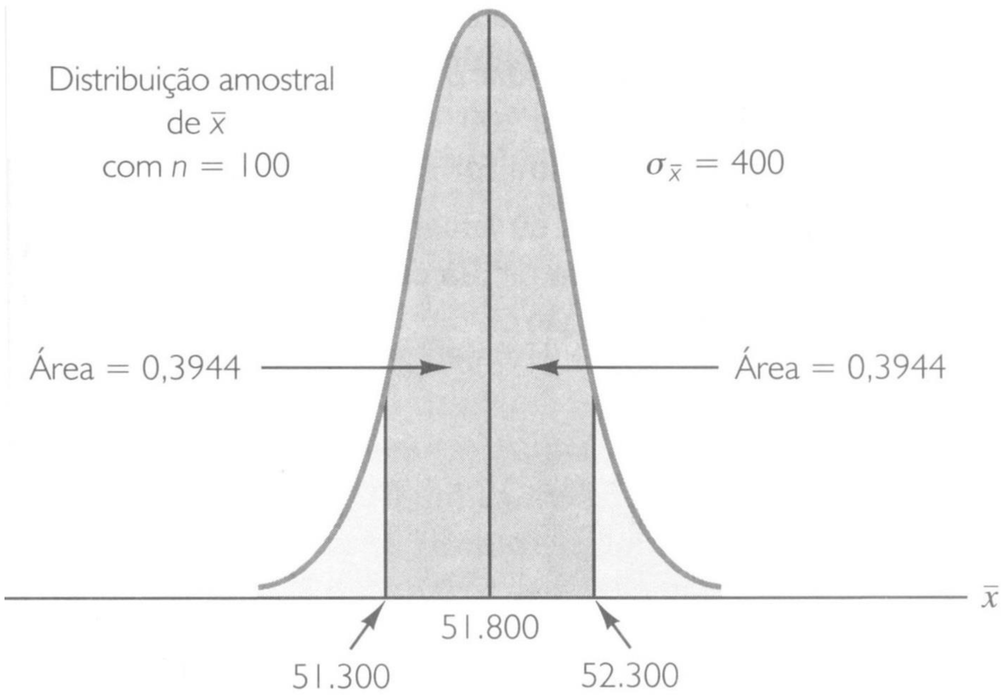


A probabilidade de uma média amostral estar dentro de US$ 500 da média da população



Relação entre tamanho da amostra e a Distribuição amostral de $\bar{x}$





DISTRIBUIÇÃO AMOSTRAL DE $\bar{p}$

A proporção amostral $\bar{p}$ é o estimador por ponto da proporção p da população. A fórmula para calcular a proporção amostral é:

$$\bar{p} = \frac{x}{n}$$

Em que

x = número de elementos contidos na amostra que possuem a característica de interesse;

n = tamanho da amostra.

A distribuição amostral de $\bar{p}$ é a distribuição de probabilidade de todos os valores possíveis da proporção amostral $\bar{p}$.

Para determinar quão próxima a proporção amostral está da proporção populacional, é preciso entender as propriedades da distribuição amostral de $\bar{p}$: o valor esperado de $\bar{p}$, o desvio padrão amostral e a forma da distribuição amostral de $\bar{p}$.

Valor esperado de $\bar{p}$*:* que é a média de todos os valores possíveis de $\bar{p}$, é igual a proporção populacional de *p*.

$$E(\bar{p}) = p$$

No caso da *EAI*, o valor esperado de $\bar{p}$ é o valor de $p = 0{,}60$.

Desvio padrão de $\bar{p}$*:* da mesma forma que procedemos para determinar o desvio padrão da média amostral, o desvio padrão da proporção amostral depende de a população ser finita ou infinita. Assim, as expressões necessárias são:

População finita: $\sigma_{\bar{p}} = \sqrt{\dfrac{N-n}{N-1}}\sqrt{\dfrac{p(1-p)}{n}}$

População infinita: $\sigma_{\bar{p}} = \sqrt{\dfrac{p(1-p)}{n}}$

Em relação ao cálculo do desvio padrão da proporção amostral, seguimos a mesma regra prática que recomendamos para o cálculo do desvio padrão da média amostral. Para populações finitas com $n/N \leq 0{,}05$ o fator de correção não precisa ser aplicado; caso a relação $n/N > 0{,}05$, o fator de correção para populações finitas deverá ser usado.
Para o caso da EAI, sabemos que a proporção de gerentes que participaram do curso de treinamento gerencial é $p = 0{,}60$; a relação $n/N = 30/2500 = 0{,}012$ nos mostra que podemos ignorar o fator de correção para populações finitas, no que resulta:

$$\sigma_{\bar{p}} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0{,}60(1-0{,}60)}{30}} 0{,}0894$$

Forma da distribuição amostral de $\bar{p}$: a proporção amostral é $\bar{p} = \dfrac{x}{n}$. Assim, o valor de $\bar{p}$ é uma variável aleatória binomial que

indica o número de elementos contidos na amostra e que possuem a característica de interesse.
Sendo *n* uma constante, a probabilidade de *x/n* é idêntica à probabilidade binomial de *x*. Isso significa que a distribuição amostral de $\bar{p}$ também é uma distribuição discreta de probabilidade.

Portanto, estamos tratando de uma distribuição binomial; ela pode ser aproximada por meio de uma distribuição normal desde que:

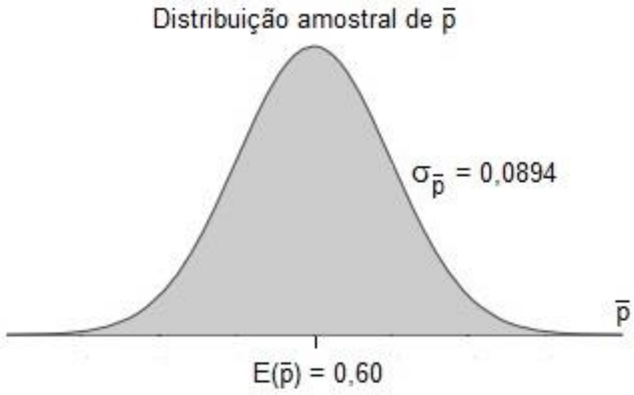
$$np \geq 5$$
$$n(1-p) \geq 5$$

Assim, pode-se escrever: "*a distribuição amostral de $\bar{p}$ pode ser aproximada por meio de uma distribuição normal sempre que a condição acima seja satisfeita*".

Lembrando o caso da *EAI*, sabemos que *p* = 0,60. Com uma amostra aleatória simples de tamanho *n* = 30, temos:
n.p = 18
n.(1-p) = 12

Assim, a distribuição amostral de $\bar{p}$ pode ser aproximada pela distribuição normal apresentada na Figura a seguir:

Distribuição amostral da proporção amostral referente à proporção de gerentes da EAI que participaram do programa de treinamento gerencial



Valor prático da Distribuição amostral de $\bar{p}$: ela pode fornecer informações probabilísticas a respeito da diferença entre a proporção amostral e a proporção da população.

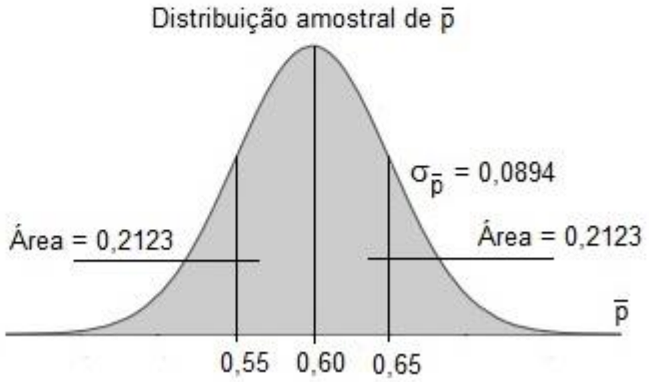
Retornemos ao exemplo da *EAI* e vamos supor que o diretor de

pessoal queira saber qual é a probabilidade de obter um valor de $\bar{p}$ que se situe no intervalo de 0,05 da proporção populacional de gerentes da *EAI* que participaram do programa de treinamento. Objetivamente o que o diretor quer saber é a probabilidade de obter uma amostra com uma proporção amostral $\bar{p}$ que se situe entre 0,55 e 0,65.
Ora, tendo em vista que podemos aproximar a distribuição amostral de $\bar{p}$ por uma distribuição normal que possua uma média igual a 0,60 e desvio padrão amostral de 0,0894, determinamos os valores da variável aleatória normal padrão (z) correspondentes a $\bar{p}$ = 0,55 e a $\bar{p}$ = 0,65.

$$z = \frac{x-\mu}{\sigma} = \frac{0,55-0,60}{0,0894} = -0,56$$

$$z = \frac{x-\mu}{\sigma} = \frac{0,65-0,60}{0,0894} = 0,56$$

Probabilidade de se obter uma proporção amostral entre 0,55 e 0,65



Entrando com os valores de z na tabela de distribuição normal padrão, encontramos os valores das respectivas áreas:
Para z = -0,56, a área entre z=-0,56 e z=0 é igual a 0,2123;
Para z = 0,56, a área entre z=0 e z=0,56 é igual a 0,2123.
Portanto, a probabilidade de selecionar uma amostra que forneça uma $\bar{p}$ dentro de 5% do valor da proporção populacional é 0,2123 + 0,2123 = 0,4246, ou seja, 42,46%.

DOIS MINUTOS PARA PENSAR: Imagine que o diretor resolva solicitar ao analista uma alternativa que produza um aumento na probabilidade de selecionar uma amostra com $\bar{p}$ que difira em 0,05 da proporção populacional. Qual a alternativa poderá produzir uma probabilidade maior que 42,46%?

OUTROS MÉTODOS DE AMOSTRAGEM

Amostragem aleatória estratificada – na amostragem aleatória estratificada, os elementos que fazem parte dos estratos devem ser o mais similares possível e devem pertencer somente a um *estrato*. Depois que os estratos são formados extrai-se uma amostra aleatória simples de cada um deles. Há fórmulas específicas disponíveis para se combinar os resultados das amostras de estrato individuais em uma estimativa do parâmetro populacional de interesse.

Amostragem por conglomerados – nesse tipo de amostragem, os elementos da população são divididos em grupos distintos denominados *conglomerados*. Extrai-se uma amostra aleatória simples de conglomerados. Todos os elementos contidos em cada conglomerado amostrado formam a amostra. Essa técnica produz bons resultados quando os elementos contidos nos conglomerados não forem similares.

Amostragem sistemática – nos casos de grandes populações, gasta-se muito tempo para selecionar uma amostra aleatória simples usando a técnica dos números aleatórios. Uma alternativa à amostragem aleatória simples é a amostragem sistemática. Vamos esclarecer o procedimento usando um exemplo.
Seja selecionar uma amostra de tamanho 50 de uma população de 5.000 elementos.
Primeiramente extrai-se um elemento em cada 5000/50 = 100 elementos. A amostra sistemática, nesse caso, tem início ao selecionar um elemento dos primeiros 100 elementos da lista da população. Feita a seleção, o seguinte será o centésimo elemento seguinte da lista, e assim sucessivamente. Vamos supor que o elemento selecionado entre os 100 primeiros da lista seja o elemento 47. A amostra de tamanho 50 será completada com os seguintes elementos:

47	147	247	347	447	547
647	747	847	947	1047	1147
1247	1347	1447	1547	1647	1747
1847	1947	2047	2147	2247	2347
2447	2547	2647	2747	2847	2947
3047	3147	3247	3347	3447	3547
3647	3747	3847	3947	4047	4147
4247	4347	4447	4547	4647	4747
4847	4947				

Amostragem de conveniência – é uma técnica de amostragem ***não probabilística***. Como o nome indica, a amostra é identificada por conveniência. Os elementos são incluídos na amostra sem probabilidades previamente especificadas ou conhecidas de eles serem selecionados. Como exemplos de uso dessa técnica temos o caso de amostras de animais selvagens capturados, grupos de voluntários para pesquisa de consumidores, professor que usa seus alunos bolsistas simplesmente porque eles estão disponíveis, etc.

Amostragem de julgamento – é uma técnica adicional de amostragem ***não probabilística***; nessa abordagem, a pessoa que conhece mais profundamente o tema do estudo escolhe os elementos que julga serem os mais representativos da população. Entretanto a qualidade dos resultados da amostra depende do julgamento da pessoa que a seleciona.

EXERCÍCIOS

1) A média de uma população é 200 e seu desvio padrão é 50. Uma amostra aleatória simples de tamanho 100 será selecionada e a média amostral será usada para estimar a média da população.
a) Qual é o valor esperado de $\bar{x}$;
b) qual é o desvio padrão de $\bar{x}$;
c) apresente a distribuição amostral de $\bar{x}$;
d) o que a distribuição amostral de $\bar{x}$ indica?

2) A média de uma população é 200 e seu desvio padrão é 50. Uma amostra aleatória simples de tamanho 100 será selecionada e a média amostral será usada para estimar a média da população.
a) qual é a probabilidade de a média da amostra estar dentro de ±5 da média da população?
b) qual é a probabilidade de a média da amostra estar dentro de ±10 da média da população?

3) Suponha que o desvio padrão da população seja 25. Calcule o erro padrão da média para tamanhos de amostra iguais a 50, 100, 150 e 200. O que se pode afirmar sobre o tamanho do erro padrão da média quando o tamanho da amostra for aumentado?

4) O custo médio do ensino nas universidades públicas norte-americanas é US$ 4.260 por ano. Use esse valor como média populacional e considere que o desvio padrão da população é US$ 900. Suponha que uma amostra aleatória de 50 universidades públicas seja selecionada.

a) apresente a distribuição amostral de $\bar{x}$ em que $\bar{x}$ é a média amostral do custo do ensino nas 50 universidades;
b) qual é a probabilidade de a amostra aleatória simples produzir uma média amostral que se situe dentro dos US$ 250 da média populacional;
c) qual é a probabilidade de a amostra aleatória simples produzir uma média amostral que se situe dentro dos US$ 100 da média populacional.

5) O custo médio anual dos seguros de automóvel é US$ 687. Use esse valor como média populacional e suponha que o desvio padrão da população seja US$ 230. Considere uma amostra de 45 apólices de seguro de automóveis.
a) apresente a distribuição amostral de $\bar{x}$ em que $\bar{x}$ é a média amostral do custo anual dos seguros de automóvel;
b) qual é a probabilidade de a média amostral estar dentro dos US$ 100 da média populacional;
c) qual é a probabilidade de a média amostral estar dentro dos US$ 25 da média populacional;
d) o que você recomendaria se uma seguradora quisesse a média amostral para estimar a média populacional dentro de ±US$ 25?

6) A proporção de uma população é 0,40. Uma amostra aleatória simples de tamanho 200 será selecionada e a proporção amostral $\bar{p}$ será usada para estimar a proporção da população.
a) qual é a probabilidade de a proporção amostral estar dentro de ±0,03 da proporção populacional;
b) qual é a probabilidade de a proporção amostral estar dentro de ±0,05 da proporção populacional?

7) A proporção populacional é 0,30. Qual é a probabilidade de a proporção amostral estar dentro de ±0,04 da proporção populacional correspondente a cada um dos seguintes tamanhos de amostra:
a) n = 100 b) n = 200 c) n = 500 d) n = 1.000
e) qual a vantagem de um tamanho de amostra maior?

8) A *Internet Express* divulgou que 56% das famílias dos EUA têm acesso à Internet. Use a proporção populacional p = 0,56 e suponha que uma amostra de 300 famílias seja selecionada.
a) apresentar a distribuição amostral de $\bar{p}$ em que $\bar{p}$ é a proporção de famílias que têm acesso à Internet;
b) qual é a probabilidade de a proporção amostral estar dentro de ±0,03 da proporção populacional;

c) responda o item (b) considerando os tamanhos de amostra 600 e 1.000.

9) Um pesquisador relata os resultados de uma pesquisa afirmando que o erro padrão da média é 20. O desvio padrão da população é 500.

a) qual é o tamanho da amostra utilizada nessa pesquisa?

b) qual a probabilidade de a estimação por ponto estar dentro de ± 25 da média da população?

Tabela de números aleatórios (cinco dígitos)

63271	59986	71744	51102	15141	80714	58683	93108	13554	79945
88547	09896	95436	79115	08303	01041	20030	63754	08459	28364
55957	57243	83865	09911	19761	66535	40102	26646	60147	15702
46276	87453	44790	67122	45573	84358	21625	16999	13385	22782
55363	07449	34835	15290	76616	67191	12777	21861	68689	03263
69393	92785	49902	58447	42048	30378	87618	26933	40640	16281
13186	29431	88190	04588	38733	81290	89541	70290	40113	08243
17726	28652	56836	78351	47327	18518	92222	55201	27340	10493
36520	64465	05550	30157	82242	29520	69753	72602	23756	54935
81628	36100	39254	56835	37636	02241	98063	89641	64953	99337
84649	48968	75215	75498	49539	74240	03466	49292	36401	45525
63291	11618	12613	75055	43915	26488	41116	64531	56827	30825
70502	53225	03655	05915	37140	57051	48393	91322	25653	06543
06426	24771	59935	49801	11082	66762	94477	02494	88215	27191
20711	55609	29430	70165	45406	78484	31639	52009	18873	96927
41990	70538	77191	25860	55204	73417	83920	69468	74972	38712
72452	36618	76298	26678	89334	33938	95567	29380	75906	91807
37042	40318	57099	10528	09925	89773	41335	96244	29002	46453
53766	52875	15987	46962	67342	77592	57651	95508	80033	69828
90585	58955	53122	16025	84299	53310	67380	84249	25348	04332
32001	96293	37203	64516	51530	37069	40261	61374	05815	06714
62606	64324	46354	72157	67248	20135	49804	09226	64419	29457
10078	28073	85389	50324	14500	15562	64165	06125	71353	77669
91561	46145	24177	15294	10061	98124	75732	00815	83452	97355
13091	98112	53959	79607	52244	63303	10413	63839	74762	50289