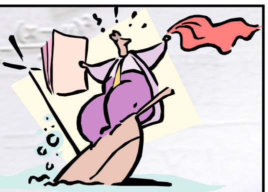
Amostragem



Prof. Lorí Viali, Dr.
viali@mat.ufrgs.br
http://www.ufrgs.br/~viali/

População

Uma coleção de todos os possíveis elementos, objetos ou medidas de interesse.

Censo

Um levantamento efetuado sobre toda uma população é denominado de **levantamento censitário** ou simplesmente **censo**.

Prof. Lorí Viali, Dr. - UFRGS - Instituto de Matemática - Departamento de Estatística

Amostra

Um subconjunto finito de uma população de interesse.

Prof. Lorí Viali, Dr. - UFRGS - Instituto de Matemática - Departamento de Estatística

Amostragem

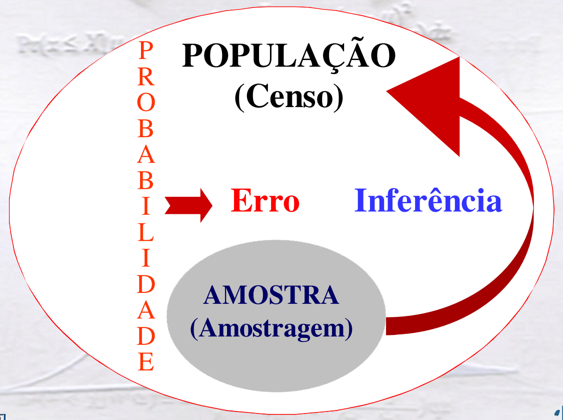
O processo de escolha de uma amostra da população é denominado de **amostragem**.

Prof. Lorí Viali, Dr. - UFRGS - Instituto de Matemática - Departamento de Estatística

Método de se inferir sobre uma população a partir do conhecimento de pelo menos uma amostra dessa população.

Prof. Lorí Viali, Dr. - UFRGS - Instituto de Matemática - Departamento de Estatística

Estudo das relações teóricas existentes entre uma população e as amostras dela extraídas.



Prof. Lorí Viali, Dr. – UFRGS – Instituto de Matemática - Departamento de Estatística

Tipos de Amostragem

- **Probabilística**
- **Não Probabilística**

Prof. Lorí Viali, Dr. – UFRGS – Instituto de Matemática - Departamento de Estatística

Amostragem Probabilística

Todos os elementos da população têm probabilidade conhecida (e diferente de zero) de fazer parte da amostra.

Métodos de Amostragem Probabilística

- Aleatória Simples
- Sistemática
- Estratificada
- Por Conglomerados

Prof. Lorí Viali, Dr. – UFRGS – Instituto de Matemática - Departamento de Estatística

Amostragem Ao Acaso (aa) ou Aleatória Simples (aas)

Uma amostra é dita "**aleatória simples**" ou "**ao acaso**" se todos os elementos da população tiverem a **mesma** probabilidade de pertencer a amostra

Prof. Lorí Viali, Dr. – UFRGS – Instituto de Matemática - Departamento de Estatística

Total de Amostras

A A S

- **Com** Reposição → $k = N^n$
- **Sem** Reposição →
  - **Não Ordenadas**: $k = \binom{N}{n}$
  - **Ordenadas**: $k = A_N^n$

Amostragem Sistemática

A unidade amostral é escolhida em intervalos pré-fixados. Assim se **N** = tamanho da população e **n** = tamanho da amostra. Então o passo ou intervalo é **k = N/n**.

Exemplo

Se N = 1000 e n = 100

Então:

$k = N/n = 1000/100 = 10.$

Sorteia-se um número entre 1 e 10. Digamos 7. Então a amostra será: 7, 17, 27, ..., 997.

Amostragem Estratificada

A população é estratificada (em grupos mutuamente exclusivos) e então uma amostra aleatória simples de cada estrato é retirada.

Amostragem por Agrupamento

Nos métodos anteriores cada observação é escolhida de forma individual. Na amostragem por agrupamento, grupos de observações são escolhidas ao acaso.

Exemplo

Considere uma população de 20 itens dividida em 5 grupos de 4 itens cada. Para escolher uma amostra de **n = 8**, escolhe-se **2 grupos**, ao invés de 8 itens individuais.

Exemplo

Grupo	Elementos
1	X_1, X_2, X_3, X_4
2	X_5, X_6, X_7, X_8
3	X_9, X_{10}, X_{11}, X_{12}
4	X_{13}, X_{14}, X_{15}, X_{16}
5	X_{17}, X_{18}, X_{19}, X_{20}

Estimador, Estimativa e Parâmetro

Uma característica da população é denominada de parâmetro.

Um estimador é uma característica da amostra.

Uma estimativa é um valor particular de um estimador.

Principais Parâmetros

μ → A MÉDIA

σ^2 → A VARIÂNCIA

σ → O DESVIO PADRÃO

π → A PROPORÇÃO

Principais Estimadores

$\overline{X}$ → A MÉDIA

S^2 → A VARIÂNCIA

S → O DESVIO PADRÃO

P → A PROPORÇÃO

Distribuições Amostrais

POPULAÇÃO

θ

$\hat{\theta}_1$ **Amostra 1**

$\hat{\theta}_2$ **Amostra 2**

....

$\hat{\theta}_k$ **Amostra k**

Distribuições Amostrais

A distribuição de probabilidade de um estimador (variável aleatória) é denominada de distribuição amostral desse estimador.

Exemplo

População $P = \{1, 2, 3, 4\}$

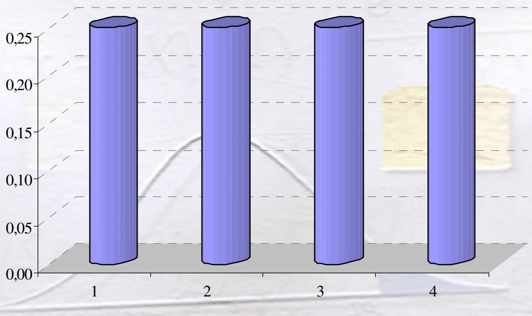
Parâmetros

$$\mu = \frac{1+2+3+4}{4} = \frac{10}{4} = 2{,}50$$

$$\sigma^2 = \frac{\sum X^2}{n} - \mu^2 = \frac{30}{4} - 2{,}50^2 = 1{,}25$$

$$\pi = \frac{0+1+0+1}{4} = \frac{2}{4} = 50\%$$

Distribuição da População



Amostras

Plano Amostral

aa = **a**o **a**caso

Método

s/r = **s**em **r**eposição

Tamanho das Amostras

n = 2

Total de Amostras

Tem-se:

$N = 4$; $n = 2$.

Então:

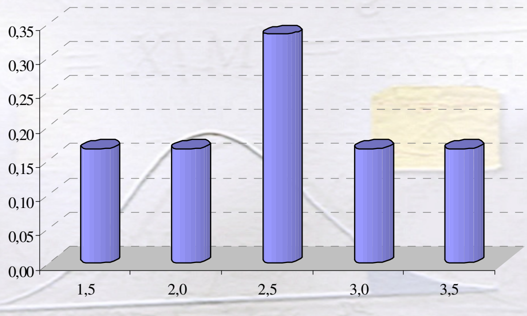
$$k = \binom{N}{n} = \binom{4}{2} = \frac{4!}{2!(4-2)!} = 6$$

	Amostras	Médias	Variâncias	Proporções
1	(1, 2)	1,5	0,5	0,5
2	(1, 3)	2,0	2,0	0,0
3	(1, 4)	2,5	4,5	0,5
4	(2, 3)	2,5	0,5	0,5
5	(2, 4)	3,0	2,0	1,0
6	(3, 4)	3,5	0,5	0,5

Distribuição Amostral da Média

$\overline{x}$	$f(\overline{x}) = P(\overline{X} = \overline{x})$
1,5	1/6
2,0	1/6
2,5	2/6
3,0	1/6
3,5	1/6
Total	**1,0**

Distribuição Amostral da Média



Prof. Lorí Viali, Dr. – UFRGS – Instituto de Matemática - Departamento de Estatística

Características da Distribuição da Média

$\overline{x}$	$f(\overline{x})$	$\overline{x}.f(\overline{x})$	$\overline{x}^2.f(\overline{x})$
1,5	1/6	1,5/6	2,25/6
2,0	1/6	2,0/6	4,00/6
2,5	2/6	5,0/6	12,50/6
3,0	1/6	3,0/6	9,00/6
3,5	1/6	3,5/6	12,25/6
Total	**1,0**	**15/6**	**40/6**

Prof. Lorí Viali, Dr. – UFRGS – Instituto de Matemática - Departamento de Estatística

Características da Distribuição da Média

$$\mu_{\overline{X}} = E(\overline{X}) = \sum \overline{x} f(\overline{x}) =$$

$$= 15/6 = 2{,}50$$

$$\sigma^2_{\overline{X}} = V(\overline{X}) = E(\overline{X}^2) - E(\overline{X})^2 =$$

$$= \frac{40}{6} - 2{,}50^2 = \frac{1{,}25}{3}$$

Prof. Lorí Viali, Dr. – UFRGS – Instituto de Matemática - Departamento de Estatística

Distribuição Amostral da Média

Características

Média $\mu_{\overline{X}} = \mathbf{E}(\overline{\mathbf{X}}) = \mu$

Erro padrão

COM Reposição $\sigma_{\overline{X}} = \dfrac{\sigma}{\sqrt{n}}$

SEM Reposição $\sigma_{\overline{X}} = \dfrac{\sigma}{\sqrt{n}}\sqrt{\dfrac{N-n}{N-1}}$

Prof. Lorí Viali, Dr. – UFRGS – Instituto de Matemática - Departamento de Estatística

Distribuição Amostral da Média

Para este exemplo, tem-se:

$$\sigma^2_{\overline{X}} = \frac{\sigma^2}{n}\frac{N-n}{N-1} = \frac{1{,}25}{2}\left(\frac{4-2}{4-1}\right) =$$

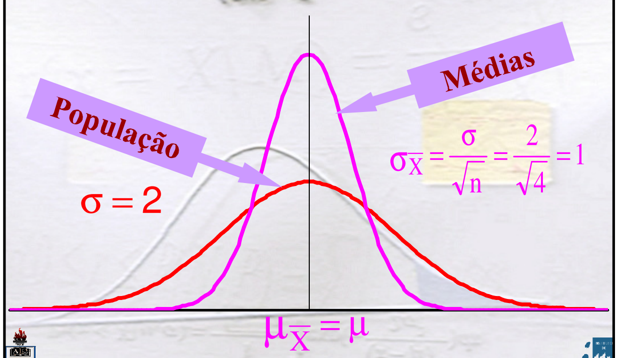
$$= \frac{1{,}25}{2}\left(\frac{2}{3}\right) = \frac{1{,}25}{3}$$

Prof. Lorí Viali, Dr. – UFRGS – Instituto de Matemática - Departamento de Estatística

Forma da Distribuição Amostral da Média

Se uma amostra aleatória de tamanho "n" for retirada de uma população X com uma distribuição N(μ; σ), então a distribuição de $\overline{X}$, média da amostra, tem uma distribuição $N(\mu, \frac{\sigma}{\sqrt{n}})$

Distribuição Amostral da Média

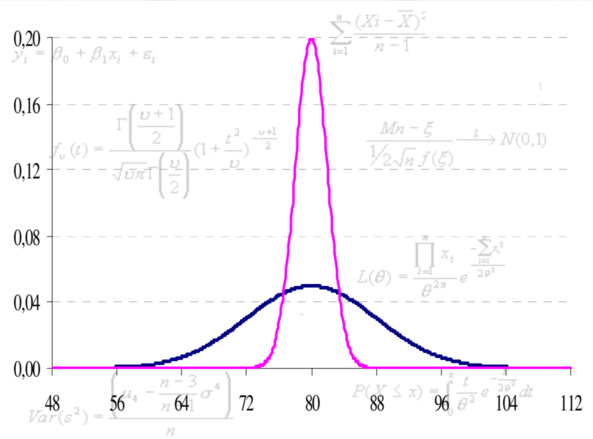


Exemplo:

Uma amostra de n = 16 elementos é retirada de uma população N(80; 8). Determine:

(a) $P(\overline{X} < 77)$

(b) $P(76 < \overline{X} < 85)$



Solução:

Tem-se: μ = 80, σ = 8

Sabe-se que:

$$\mu_{\overline{X}} = 80 \text{ e}$$

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} = \frac{8}{\sqrt{16}} = 2$$

Então:

(a) $P(\overline{X} < 77) =$

$$= P\left(\frac{\overline{X} - \mu_{\overline{X}}}{\sigma_{\overline{X}}} < \frac{77 - 80}{2}\right) =$$

$$= P(Z < -1{,}50) = \Phi(-1{,}50) =$$

$$= 0{,}0668 = 6{,}68\%$$

(b) $P(76 < \overline{X} < 85) =$

$$= P\left(\frac{76-80}{2} < \frac{\overline{X}-\mu_{\overline{X}}}{\sigma_{\overline{X}}} < \frac{85-80}{2}\right) =$$

$$= P(-2 < Z < 2{,}5) =$$

$$= \Phi(2{,}50) - \Phi(2{,}00) =$$

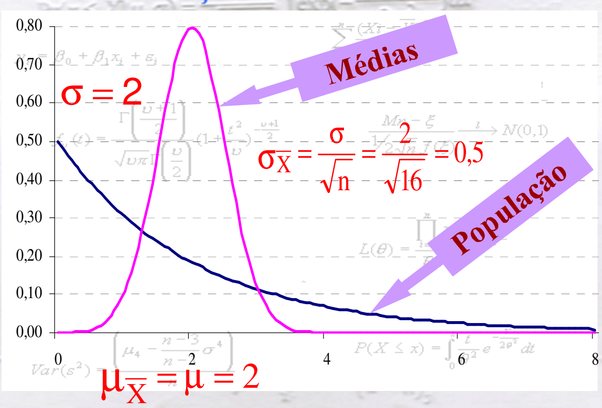
$$= 99{,}38\% - 2{,}28\% = 97{,}10\%$$

Forma da Distribuição Amostral da Média

Se uma amostra aleatória de tamanho "**n > 30**" for retirada de uma população com **qualquer** distribuição de média μ e desvio padrão σ, então a distribuição de $\overline{X}$, média da amostra, tem uma distribuição aproximadamente $N(\mu, \frac{\sigma}{\sqrt{n}})$

Prof. Lorí Viali, Dr. - UFRGS - Instituto de Matemática - Departamento de Estatística

Distribuição Amostral da Média



Exemplo:

Uma amostra de "n" elementos é retirada de uma população N(80; 4). Determine "n" de forma que:

$$P(\overline{X} < 79) = 1{,}50\%$$

Prof. Lorí Viali, Dr. - UFRGS - Instituto de Matemática - Departamento de Estatística

Solução:

Tem-se: $\mu = 80$, $\sigma = 4$

Sabe-se que:

$$\mu_{\overline{X}} = 80 \ \text{ e}$$

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} = \frac{4}{\sqrt{n}}$$

Prof. Lorí Viali, Dr. - UFRGS - Instituto de Matemática - Departamento de Estatística

Então:

$$P(\overline{X} < 79) =$$

$$= P\left(\frac{\overline{X}-\mu_{\overline{X}}}{\sigma_{\overline{X}}} < \frac{79-80}{\frac{4}{\sqrt{n}}}\right) =$$

$$= P\left(Z < -\frac{\sqrt{n}}{4}\right) = \Phi\left(-\frac{\sqrt{n}}{4}\right) = 1{,}50\%$$

Prof. Lorí Viali, Dr. - UFRGS - Instituto de Matemática - Departamento de Estatística

$$-\frac{\sqrt{n}}{4} = -2{,}17$$

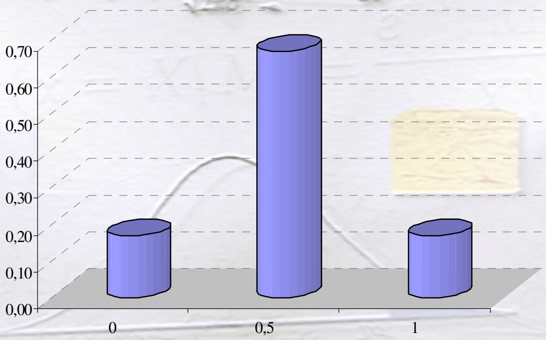
$$\sqrt{n} = 2{,}17.4 = 8{,}68$$

$$n \geq (8{,}68)^2 \cong 76$$

Distribuição Amostral da Proporção

p	**f(p)**
0,0	1/6
0,5	3/6
1,0	1/6
Total	**1,0**

Distribuição Amostral da Proporção



Características da Distribuição da Proporção

p	f(p)	p.f(p)	p^2.f(p)
0,0	1/6	0/6	0/6
0,5	4/6	2/6	1/6
1,0	1/6	1/6	1/6
Total	**1,0**	**3/6**	**2/6**

Características da Distribuição da Proporção

$$\mu_P = E(P) = \sum p.f(p) =$$

$$= 3/6 = 0{,}50 = 50\%$$

$$\sigma_P^2 = V(P) = E(P^2) - E(P)^2 =$$

$$= \frac{2}{6} - \left(\frac{3}{6}\right)^2 = \frac{1}{12}$$

Distribuição Amostral da Proporção

Características

Média → $\mu_P = E(P) = \pi$

Erro padrão →

COM Reposição $\sigma_P = \sqrt{\frac{\pi(1-\pi)}{n}}$

SEM Reposição $\sigma_P = \sqrt{\frac{\pi(1-\pi)}{n}}\sqrt{\frac{N-n}{N-1}}$

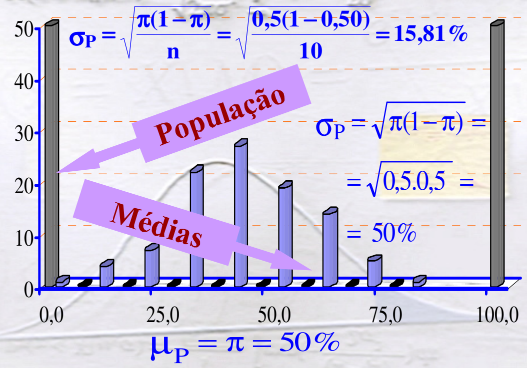
Distribuição Amostral da Proporção

Para este exemplo, tem-se:

$$\sigma_P^2 = \frac{\pi(1-\pi)}{n}\frac{N-n}{N-1} = \frac{0{,}5.0{,}5}{2}\left(\frac{4-2}{4-1}\right) =$$

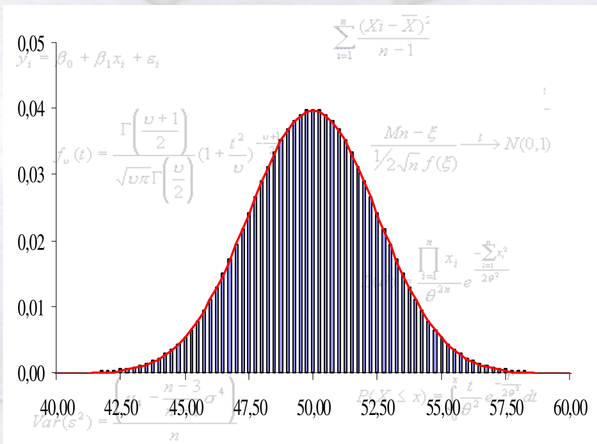
$$= \frac{0{,}25}{2}\left(\frac{2}{3}\right) = \frac{0{,}25}{3} = \frac{1}{12}$$

Distribuição Amostral da Proporção



Forma da Distribuição Amostral da Proporção

Se uma amostra aleatória de tamanho "**n > 100**" for retirada de uma população com proporção **π**, então a distribuição de **P**, **proporção na amostra**, tem uma distribuição aproximadamente $N(\pi, \sqrt{\frac{\pi(1-\pi)}{n}})$



Exemplo:

Uma amostra de n = 400 eleitores é retirada da população que prefere o candidato Zigoto com π = 50% Determine:

(a) $P(47\% < P < 54\%)$

(b) $P(P > 56\%)$

Solução:

Tem-se: π = 50%

Sabe-se que: $\mu_P = \pi = 50\%$

$$\sigma_P = \sqrt{\frac{\pi(1-\pi)}{n}} =$$

$$= \sqrt{\frac{0{,}45(1-0{,}45)}{400}} =$$

$$= 0{,}025 = 2{,}50\%$$

Então:

(a) $P(47 < P < 54) =$

$$= P\left(\frac{47\% - 50\%}{2{,}5\%} < \frac{P - \mu_P}{\sigma_P} < \frac{54\% - 50\%}{2{,}5\%}\right) =$$

$$= P(-1{,}20 < Z < 1{,}60) =$$

$$= \Phi(1{,}60) - \Phi(-1{,}20) = 94{,}52\% - 11{,}51\% =$$

$$= 83{,}01\%$$

(b) $P(P > 56\%) =$

$$= P\left(\frac{P - \mu_P}{\sigma_P} > \frac{56\% - 50\%}{2{,}50\%}\right)$$

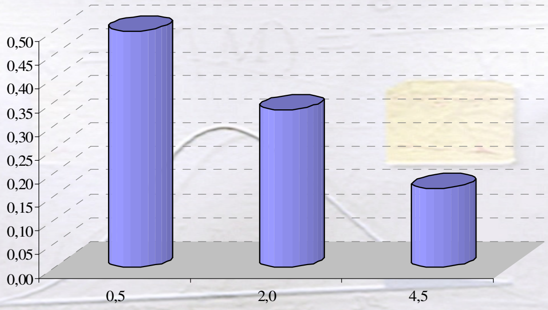
$$= P(Z > 2{,}40) = 1 - \Phi(2{,}40) =$$

$$= \Phi(-2{,}40) = 0{,}82\%$$

Distribuição Amostral da Variância

s^2	$f(s^2)$
0,5	3/6
2,0	2/6
4,5	1/6
Total	**1,0**

Distribuição Amostral da Variância



Distribuição Amostral da Variância

s^2	$f(s^2)$	$s^2.f(s^2)$	$(s^2)^2.f(s^2)$
0,5	3/6	1,5/6	0,75/6
2,0	2/6	4,0/6	8,00/6
4,5	1/6	4,5/6	20,25/6
Total	**1,0**	**10/6**	**29/6**

Distribuição Amostral da Variância

$$\mu_{S^2} = E(S^2) = \sum s^2 f(s^2) =$$

$$= \frac{5}{3} = 1{,}67$$

$$\sigma^2_{S^2} = V(S^2) = E[(S^2)^2] - E(S^2)^2 =$$

$$= \frac{29}{6} - \left(\frac{5}{3}\right)^2 = \frac{87 - 50}{18} = \frac{37}{18} = 2{,}06$$

Distribuição Amostral da Variância

Características

Amostragem **com** reposição

Média $\mu_{S^2} = E(S^2) = \sigma^2$

Erro padrão $\sigma_{S^2} = \sqrt{\dfrac{2\sigma^4}{n-1}} = \sigma^2\sqrt{\dfrac{2}{n-1}}$

Forma da Distribuição Amostral da Variância

Se uma amostra aleatória de tamanho "**n**" (grande) for retirada de uma população com variância $\boldsymbol{\sigma^2}$, então a distribuição de $\mathbf{S^2}$, **variância da amostra**, tem uma distribuição aproximadamente χ^2 com "n-1" g.l., a menos de uma constante.

Distribuição Amostral da Variância

Isto é:

$$S^2 = \frac{\sigma^2}{n-1}\chi^2_{n-1}$$

Este resultado é conhecido como Teorema de Fisher

Exemplo:

Uma amostra de **n = 81** elementos é retirada de uma população com variância $\boldsymbol{\sigma^2 = 10}$. Determine a probabilidade de que $P(S^2 > 15)$.

Solução:

Tem-se:

$n = 81$

$\sigma^2 = 10$

Sabe-se que:

$$S^2 = \frac{\sigma^2}{n-1}\chi^2_{n-1}$$

Então:

$$P(S^2 > 15) = P[\frac{\sigma^2}{(n-1)}\chi^2_{n-1} > 15] =$$

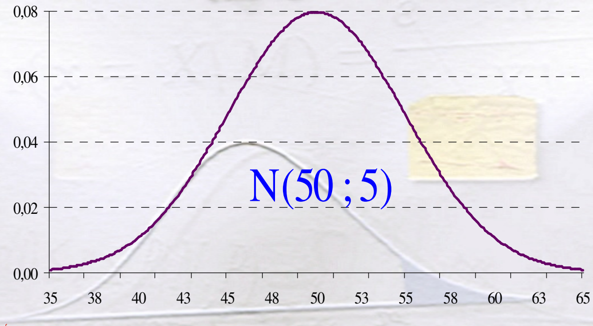
$$= P[\chi^2_{n-1} > \frac{15.(n-1)}{\sigma^2}] =$$

$$= P(\chi^2_{80} > \frac{15.80)}{10}) = P(\chi^2_{80} > \frac{15.80)}{10}) =$$

$$P(\chi^2_{80} > 120) = 0{,}25\,\%$$

Simulações

População Amostrada



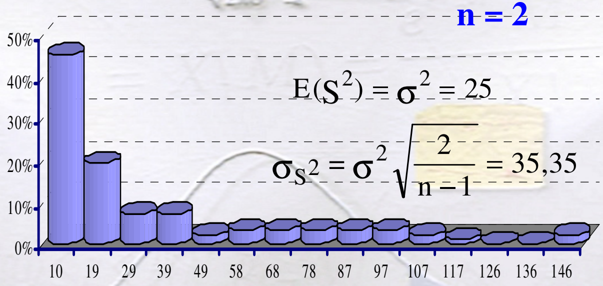
Prof. Lorí Viali, Dr. - UFRGS - Instituto de Matemática - Departamento de Estatística

Distribuição Amostral da Variância

n = 2

$E(S^2) = \sigma^2 = 25$

$\sigma_{S^2} = \sigma^2 \sqrt{\frac{2}{n-1}} = 35{,}35$



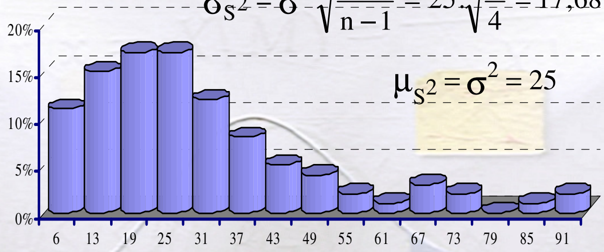
Mínimo	Máximo	Média	Desvio (Erro) Padrão
0,0085	110,2515	22,0809	25,76778

Distribuição Amostral da Variância

n = 5

$\sigma_{S^2} = \sigma^2 \sqrt{\frac{2}{n-1}} = 25.\sqrt{\frac{2}{4}} = 17{,}68$

$\mu_{S^2} = \sigma^2 = 25$



Mínimo	Máximo	Média	Desvio (Erro) Padrão
3,54	113,22	26,80	20,37

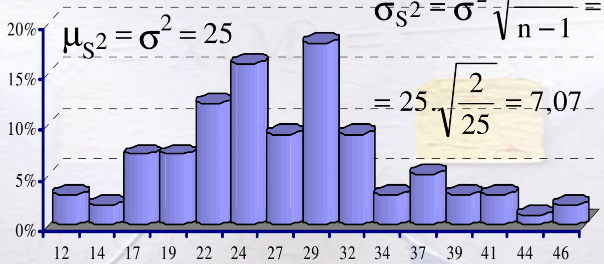
Distribuição Amostral da Variância

n = 25

$\mu_{S^2} = \sigma^2 = 25$

$\sigma_{S^2} = \sigma^2 \sqrt{\frac{2}{n-1}} =$

$= 25.\sqrt{\frac{2}{25}} = 7{,}07$



Mínimo	Máximo	Média	Desvio (Erro) Padrão
12,94	39,90	25,66	6,28