Técnicas de Amostragem - Aula 04

Exercícios, Amostragem Aleatória Simples Sem Reposição (AASs) I

Kaique Matias de Andrade Roberto

Ciências Atuariais

HECSA - Escola de Negócios

FIAM-FAAM-FMU

$\boxed{2.3}$ $\bar{F}$ é a média amostral da variável F

$$D = \begin{array}{c} \begin{array}{ccc} i & F_i & T_i \end{array} \\ \begin{pmatrix} 1 & 12 & 1 \\ 2 & 30 & 3 \\ 3 & 18 & 2 \end{pmatrix} \end{array}$$

$$\bar{F}(12) \quad \frac{12+30}{2} = \frac{42}{2} = 21$$

$\boxed{1.5}$

$$S_3(U) = \{ (111); (112); (113); \ldots \}$$

Conteúdo

Conceitos que aprendemos em Aulas anteriores

Conceitos que aprendemos em Aulas anteriores

Na aula passada nós lidamos com os aspectos da AASc:

- definição da AASc;
- propriedades das principais estatísticas;
- normalidade e intervalo de confiança;
- tamanho da amostra.

Conceitos que aprendemos em Aulas anteriores

Agora vamos resolver alguns exercícios para fixar os conceitos da AASc.

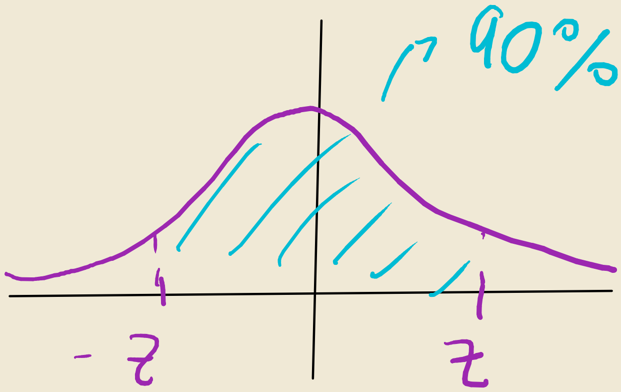
Exercício 3.4 (A). Seja $X \sim N(\mu, \sigma^2)$. Calcule:

a - $P(-\sigma + \mu < Z < \sigma + \mu)$;

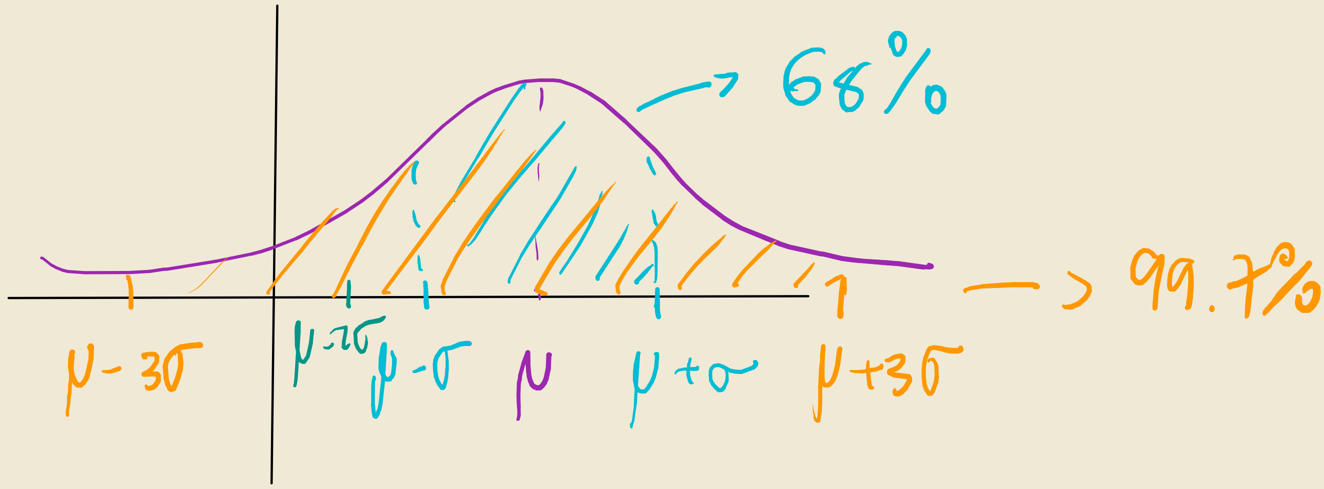
b - $P(-2\sigma + \mu < Z < 2\sigma + \mu)$;

c - $P(-3\sigma + \mu < Z < 3\sigma + \mu)$;

d - $P(-4\sigma + \mu < Z < 4\sigma + \mu)$.



a -



$$P(\mu - \sigma < X < \mu + \sigma) =$$

$$P\left(\frac{\mu - \sigma - \mu}{\sigma} < \frac{X - \mu}{\sigma} < \frac{\mu + \sigma - \mu}{\sigma}\right) =$$

$$P(-1 < Z < 1)$$

$$P(-1 < Z < 1) = P(Z < 1) - P(Z < -1)$$

$$= 0.8413 - 0.1587 = 0.6827$$

$$P(\mu - 2\sigma < X < \mu + 2\sigma) = P(-2 < Z < 2) =$$

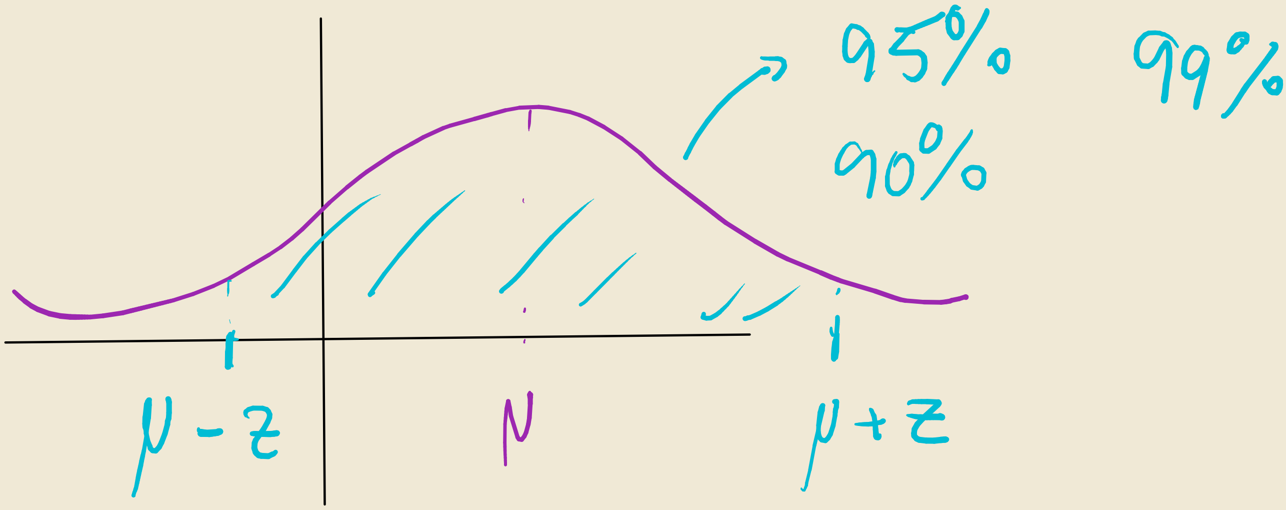
$$P(Z < 2) - P(Z < -2) =$$

$$0.9772 - 0.0228 = 0.9744$$

$$P(\mu - 3\sigma < X < \mu + 3\sigma) = P(-3 < Z < 3) =$$

$$P(Z < 3) - P(Z < -3) =$$

$$0.9987 - 0.0013 = 0.9974$$



$$P(\mu - z < X < \mu + z) =$$

$$P\left(\frac{\mu - z - \mu}{\sigma} < \frac{X - \mu}{\sigma} < \frac{\mu + z - \mu}{\sigma}\right) =$$

$$P\left(\frac{-z}{\sigma} < Z < \frac{z}{\sigma}\right) =$$

$$P\left(Z < -\frac{z}{\sigma}\right) - P\left(Z < \frac{z}{\sigma}\right)$$

$\boxed{\text{Vamos resolver p/ } Z \sim N(0,1)}$

$$P(-z < Z < z) = \begin{matrix} 0.90 & 0.80 \\ 0.95 & 0.75 \\ 0.99 & 0.45 \\ & \vdots \end{matrix}$$

$\hookrightarrow P(Z < z) - P(Z < -z)$

z	$P(Z<z) - P(Z<-z)$
1.5	$0.9332 - 0.0668 = 0.8664$
1.6	$0.9452 - 0.0548 = 0.8904$
1.65	$0.9505 - 0.0505 = 0.9$
1.9	$0.9713 - 0.0287 = 0.9426$
1.93	$0.9732 - 0.0268 = 0.9464$
1.96	$0.9750 - 0.0250 = 0.95$
2.34	$0.9901 - 0.0099 = 0.9802$

P/ $X \sim N(\mu, \sigma^2)$

$= z_\gamma$

1.96   1.96

$$\frac{z}{\sigma} = 1.65 \Rightarrow z = 1.65\sigma$$

0.95

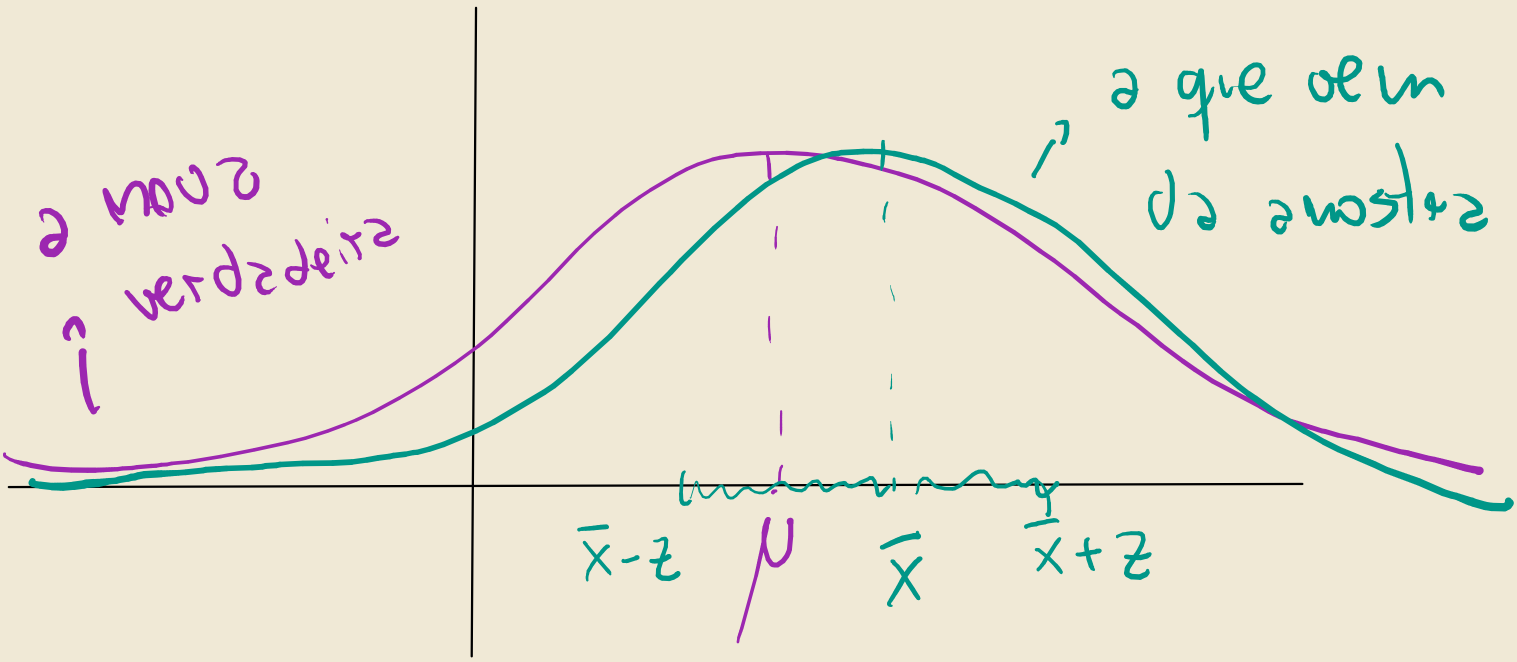
$$P(\mu - z < X < \mu + z) = 0.9$$

Exercício 3.7 (A). Uma máquina enche pacotes de café com uma variância igual a $100g$. Ela estava regulada para encher os pacotes com $500g$, em média. Agora, ela se desregulou, e queremos saber qual a nova média μ. Uma amostra de 25 pacotes apresentou uma média igual a $485g$. Vamos construir um $\boxed{\text{intervalo de confiança com 95\% de confiança para } \mu}$.

$X \sim$ peso dos pacotes de café

$X \sim N(500, 100)$ $\mu = 500$, $\sigma^2 = 100$, $\sigma = 10$

↳ a antiga



gostaríamos que $\mu \in (\bar{x} - z ; \bar{x} + z)$

Problemas:

1 - Não sabemos quem é o μ "verdadeiro"

2 - Não dá p/ ter certeza de que $\mu \in (\bar{x} - z, \bar{x} + z)$

Solução: Estabelecemos um nível de confiança γ para

$$\mu \in (\bar{x} - z ; \bar{x} - z)$$

$$P(-z < \bar{x} - \mu < z) = \gamma$$

$$P(-z < \bar{x} - \mu < z) = 0.95$$

$$P\left(\frac{-z}{\sigma_{antigo}} < \frac{\bar{x} - \mu^{antigo}}{\sigma_{antigo}} < \frac{+z}{\sigma_{antigo}}\right) = 0.95$$

$$P\left(\frac{-z}{\sigma} < Z < \frac{z}{\sigma}\right) = 0.95$$

$$\boxed{P\left(-\frac{z}{\sigma} < Z < \frac{z}{\sigma}\right) = 0.95}$$

$$\frac{z}{\sigma} = 1.96 \Rightarrow \frac{z}{10} = 19.6$$

$$\Rightarrow z = 19.6$$

Com 95% de confiança

$$\mu \in (\bar{x} - z ; \bar{x} + z)$$

$$(485 - 19.6 ; 485 + 19.6)$$

$$\boxed{\frac{z}{\sigma} = z_{\gamma}}$$

Exercício 3.8 (A). Considere a população de moradores de um condomínio ($N = 540$). Deseja-se estimar a idade média dos condôminos. Com base em pesquisas passadas, pode-se obter a estimativa para $\boxed{\sigma^2 \text{ de } 463.32.}$ Suponha que será retirada da população uma amostra segundo AASc. Admitindo que a diferença entre a média amostral e a verdadeira média populacional seja, no máximo, de 4 anos, com um nível de confiança de 95%, determine o tamanho da amostra a ser coletada.

$$P\left(|\bar{X} - \mu| < 4 \right) = 0.95$$

$$P\left(-4 < \bar{X} - \mu < 4 \right) = 0.95$$

$$P\left(\frac{-4}{\sqrt{\sigma^2/n}} < \frac{\bar{X} - \mu}{\sqrt{\sigma^2/n}} < \frac{4}{\sqrt{\sigma^2/n}} \right) = 0.95$$

$$P\left(\frac{-4}{\sqrt{\sigma^2/n}} < Z < \frac{4}{\sqrt{\sigma^2/n}} \right) = 0.95$$

$$\frac{4}{\sqrt{\sigma^2/n}} = 1.96 \Rightarrow$$

$$4 = 1.96\sqrt{\frac{\sigma^2}{n}} \Rightarrow$$

$$16 = \frac{1.96^2 \sigma^2}{n} \Rightarrow$$

$$n = \frac{1.96^2 \cdot \sigma^2}{16} = \frac{1.96^2 \cdot 463.32}{16} = 112$$

Com 95% de confiança se o tamanho da amostra for maior ou igual a 112 então

$$|\bar{x} - \mu| < 4$$

Exercício 3.9 (A). Suponha que $p = 30\%$ dos estudantes de uma escola sejam mulheres. Colhemos uma AAS de $n = 10$ estudantes e calculamos

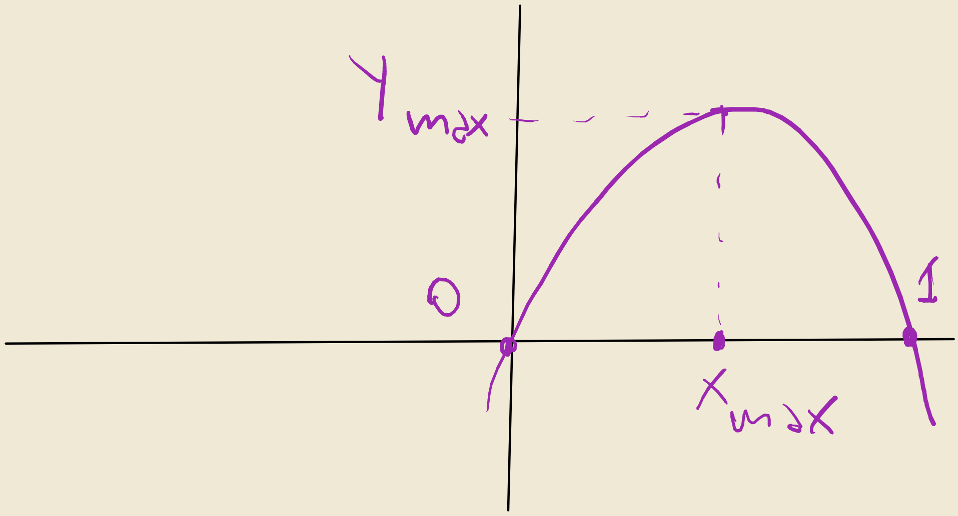
$$\hat{p} = \text{ proporção de mulheres na amostra.}$$

Qual a probabilidade de que $\hat{p}$ difira de p em menos de $0,01$?

$$Var(B(n,p)) = np(1-p) \leq n\frac{1}{4}$$

—— n ——

$$F(x) = x(1-x) = x - x^2$$



$$x_{max} = -\frac{b}{2a} = \frac{1}{2}$$

$$Y_{max} = \frac{1}{4}$$

$$P(|\hat{p}-p| < 0.01) = ?$$

$$P(-0.01 < \hat{p}-p < 0.01) =$$

$$P\left(\frac{-0.01}{\sqrt{\sigma^2/n}} < \frac{\hat{p}-p}{\sqrt{\sigma^2/n}} < \frac{0.01}{\sqrt{\sigma^2/n}}\right) =$$

$$P\left(\frac{-0.01}{\sqrt{\sigma^2/n}} < Z < \frac{0.01}{\sqrt{\sigma^2/n}}\right) =$$

$$P\left(\frac{-0.01}{\sqrt{\frac{1}{40}}} < Z < \frac{0.01}{\sqrt{\frac{1}{40}}}\right) =$$

$$P(-0.0625 < Z < 0.0625) =$$

$$P(Z<0.0625) - P(Z<-0.0625) =$$

$$0.5239 - 0.4762 = 0.0478$$

$$\approx 5\%$$

$$P(|\hat{p} - p| < 0.01) = 5\%$$

Exercício 4.1 (A). Em uma população com $N = 6$, tem-se $\boldsymbol{D} = (8, 2, 2, 11, 4, 7)$. Um plano AASc de tamanho 2 foi adotado.

1. Encontre a distribuição de $\overline{y}$ e mostre que $E[\overline{y}] = \mu$.
2. Encontre $\text{Var}(\overline{y})$.
3. Suponha que uma AASc de tamanho $n = 10$ foi retirada da população com $\overline{y} = 5.435$ e $s^2 = 3.6$. Encontre um intervalo de confiança para μ com $\alpha = 0.02$.

$\gamma = 1 - \alpha$

3 $\boxed{\text{Caso Normal}}$

$$P(|\bar{x} - \mu| < z) = 0.98$$

$$\frac{z}{s/\sqrt{n}} = 2.33 \Rightarrow$$

$$z = 2.33\sqrt{\frac{3.6}{10}} \cong 1.4$$

$$\mu \in (5.435 - 1.4 ; 5.435 + 1.4)$$

$$\mu \in (4.035 ; 6.835).$$

1

$$D = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 8 & 2 & 2 & 11 & 4 & 7 \end{pmatrix} \quad \begin{array}{l} \mu = 34/6 \\ \sigma^2 = \ldots \\ S^2 = \ldots \end{array}$$

$$AAs(2) = \{ (11); (12); (13); \ldots ; (66) \}$$

$$\bar{y}(11) = \frac{8+8}{2} = 8$$

$$\bar{y}(22) = \frac{8+2}{2} = 5$$

(Continua nas próximas aulas ...)

Amostragem Aleatória Simples

Amostragem Aleatória Simples

A amostragem aleatória simples (AAS) é o método mais simples e mais importante para a seleção de uma amostra.

Amostragem Aleatória Simples

Além de servir como plano próprio, o seu procedimento é usado de modo repetido em procedimento de múltiplos estágios.

Amostragem Aleatória Simples

A principal caracterização para o uso do plano AAS é a existência de um sistema de referências completo, descrevendo cada uma das unidades elementares.

Amostragem Aleatória Simples

Deste modo tem-se bem listado o universo

$$\mathcal{U} = \{1, 2, ..., N\}.$$

O plano é descrito do seguinte modo:

Amostragem Aleatória Simples

- Utilizando algum procedimento aleatório (tabela de números aleatórios, urna, etc) sorteia-se com igual probabilidade um elemento da população $\mathcal{U}$;

Amostragem Aleatória Simples

- repete-se o processo anterior até que sejam sorteados n unidades, tendo sido este número pré-fixado anteriormente;

Amostragem Aleatória Simples

- caso seja permitido o sorteio de uma unidade mais de uma vez, tem-se o processo AAS com reposição (AASc). Quando o elemento sorteado é removido de $\mathcal{U}$ antes do sorteio próximo, tem-se o plano AAS sem reposição (AASs).

Amostragem Aleatória Simples

Do ponto de vista prático, o plano AASs é muito mais interessante, pois vai de encontro ao princípio intuitivo que "não se ganha mais informação se uma mesma unidade aparece mais de uma vez na amostra".

Amostragem Aleatória Simples

Por outro lado, o plano AASc, introduz vantagens matemáticas e estatísticas, como a independência entre as unidades sorteadas, que facilita em muito a determinação das propriedades estatísticas dos estimadores das quantidades populacionais de interesse.

Amostragem Aleatória Simples

Basta observar na maioria dos assuntos tratados em livros de inferência há imposição de que as unidades que fazem parte da amostra sejam independentes.

Amostragem Aleatória Simples

Também vamos considerar inferências para os seguintes parâmetros de interesse: considere para cada unidade i, uma característica populacional unidimensional de interesse, Y_i, $i \in \mathcal{U}$. Vamos considerar:

Amostragem Aleatória Simples

- total populacional,

$$\tau = \sum_{i=1}^{N} Y_i;$$

Amostragem Aleatória Simples

- média populacional,

$$\mu = \overline{Y} = \frac{1}{N}\sum_{i=1}^{N} Y_i;$$

Amostragem Aleatória Simples

- variância populacional, representada por

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(Y_i - \mu)^2 \text{ e } S^2 = \frac{1}{N-1}\sum_{i=1}^{N}(Y_i - \mu)^2.$$

Amostragem Aleatória Simples Sem Reposição (AASs)

Amostragem Aleatória Simples Sem Reposição (AASs)

A AASs opera da seguinte forma:

Amostragem Aleatória Simples Sem Reposição (AASs)

- A população está numerada de 1 a N, de acordo com o sistema de referências, ou seja,
$$\mathcal{U} = \{1, 2, ..., N\}.$$

Amostragem Aleatória Simples Sem Reposição (AASs)

- Utilizando um software ou tabela de números aleatórios, sorteia-se, com igual probabilidade, uma das N unidades da população;

Amostragem Aleatória Simples Sem Reposição (AASs)

- sorteia-se um elemento seguinte, com o elemento anterior sendo retirado da população;

Amostragem Aleatória Simples Sem Reposição (AASs)

- repete-se esse procedimento até que n unidades tenham sido sorteadas.

Amostragem Aleatória Simples Sem Reposição (AASs)

Teorema 3.1

Para o plano amostral AASs, a variável f_i, número de vezes que a unidade i aparece na amostra segue uma distribuição Bernoulli $f_i \sim Ber\left(\frac{n}{N}\right)$ e que satisfaz

$$P(f_i = 1) = \frac{n}{N} \text{ e } P(f_i = 0) = 1 - \frac{n}{N},$$

de modo que

$$E[f_i] = \frac{n}{N}, \text{Var}[f_i] = \frac{n}{N}\left(1 - \frac{1}{N}\right)$$

$$\text{Cov}[f_i, f_j] = -\frac{n}{N^2}\frac{N-n}{N-1} \text{ para } i \neq j.$$

Amostragem Aleatória Simples Sem Reposição (AASs)

Convém ressaltar ainda a similaridade entre muitos dos resultados que os dois planos apresentam, e que embora as fórmulas sejam diferentes, são próximas quando N, o tamanho da população, tende a ser grande quando comparado com o tamanho da amostra.

Teorema 3.2

A estatística $t(\boldsymbol{s})$, total da amostra, definida por

$$t(\boldsymbol{s}) = \sum_{i \in \boldsymbol{s}} Y_i$$

tem, para o plano AASs, as seguintes propriedades:

$$E[t] = n\mu \text{ e } \mathrm{Var}[t] = n(1 - f)\sigma^2,$$

onde $f = n/N$ é denominada fração amostral.

Corolário 3.3

Com relação à AASs, um estimador não-viesado do total populacional é

$$T(\boldsymbol{s}) = N\overline{y} = \frac{N}{n} t(\boldsymbol{s}),$$

cuja variância amostral é dada por

$$\mathrm{Var}[T] = N^2(1 - f)\frac{S^2}{n}.$$

Corolário 3.4

Com relação à AASs, a média amostral

$$\overline{y} = \frac{1}{n}\sum_{i \in \boldsymbol{s}}^{n} Y_i = \frac{t(\boldsymbol{s})}{n}$$

é um estimador não-viesado da média populacional, com variância amostral dada por

$$\mathrm{Var}[\overline{y}] = (1 - f)\frac{S^2}{n}.$$

Amostragem Aleatória Simples Sem Reposição (AASs)

Teorema 3.5

A variância amostral

$$s^2 = \frac{1}{n-1} \sum_{i \in \boldsymbol{s}} (Y_i - \overline{y})^2$$

é um estimador não viesado da variância populacional S^2 para o planejamento amostral AASs.

Amostragem Aleatória Simples Sem Reposição (AASs)

Exemplo 3.6

Considere novamente os dados na planilha "aula-02-exemplo", e considere a variável renda familiar, onde o universo é $\mathcal{U}$ e o parâmetro populacional é $\boldsymbol{D} = (12, 30, 18)$. Verifique o comportamento das estatísticas $\overline{y}$ e s^2 com relação as funções paramétricas μ e σ^2 de $\boldsymbol{D}$ para o plano amostral AASs com $n = 2$.

$$D = \begin{pmatrix} 1 & 12 \\ 2 & 30 \\ 3 & 18 \end{pmatrix} \qquad \mu = 20 \ ; \ \sigma^2 = \frac{166}{3} = 56$$

$$S^2 = \frac{166}{2} = 74$$

$$AAS_s(2) =$$

$$\{ (12) ; (13) ; (21) ; (23) ; (31) ; (32) \}$$

s	(12)	(13)	(21)	(23)	(31)	(32)
$\overline{Y}$	21	15	21	24	15	24
s^2	81	117	81	36	117	36

h	15	21	24
$P(\overline{Y} = h)$	$2/6$	$2/6$	$2/6$

h	36	81	117
$P(s^2 = h)$	$2/6$	$2/6$	$2/6$

h	15	21	24
$P(\bar{y}=h)$	$2/6$	$2/6$	$2/6$

$$E[\bar{y}] = \frac{15\cdot 2}{6} + \frac{21\cdot 2}{6} + \frac{24\cdot 2}{6}$$

$$= \frac{1}{3}(15+21+24) = \frac{60}{3} = 20 = \mu$$

h	36	81	117
$P(s^2=h)$	$2/6$	$2/6$	$2/6$

$$E[s^2] = 36\cdot\frac{2}{6} + 81\cdot\frac{2}{6} + 117\cdot\frac{2}{6}$$

$$= \frac{1}{3}[36+81+117 = \frac{234}{3} = \ldots = S^2$$

Comentários Finais

Comentários Finais

Em resumo, na aula de hoje nós resolvemos alguns exercícios e lidamos com os seguintes aspectos da AASs:

- exercícios envolvendo AASc;
- definição da AASs;
- propriedades das principais estatísticas.

Comentários Finais

Na próxima aula nós resolvemos mais exercícios e lidaremos com os outros aspectos da AASs:

- normalidade e intervalo de confiança;
- tamanho da amostra;
- estimação da proporção;
- comparação entre AASc e AASs.

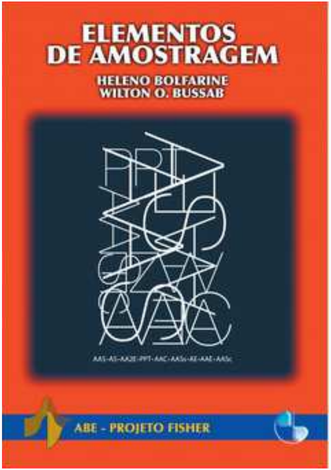
Comentários Finais

ATIVIDADE PARA ENTREGAR (E COMPOR A NOTA N1)

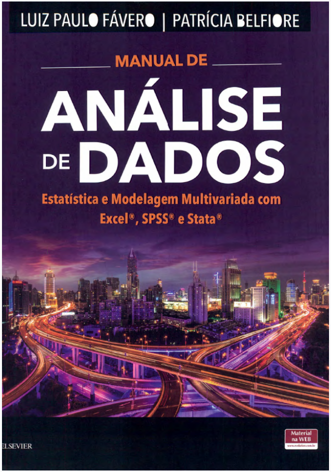
Resolva em grupos de até 4 integrantes os Exercícios 4.1-4.3.

Referências

Referências



Referências



Bons Estudos!