Técnicas de Amostragem - Aula 07

Comentários Pós N1 - Exercícios

Kaique Matias de Andrade Roberto

Ciências Atuariais

HECSA - Escola de Negócios

FIAM-FAAM-FMU

Conteúdo

Conceitos que aprendemos em Aulas anteriores

Conceitos que aprendemos em Aulas anteriores

- definição da AASc e AASs;
- propriedades das principais estatísticas;
- normalidade e intervalo de confiança;
- tamanho da amostra;
- fizemos uma breve introdução aos geradores de números aleatórios.

População ("Espaço de Eventos") - é um conjunto qualquer não-vazio.

$\mathcal{U} = \{1, \dots, N\}$

Espaço Amostral ("Espaço de Probabilidades")

$$S \subseteq \wp(\mathcal{U})$$

$\mathcal{U} = \{1, 2, 3\} \rightarrow$ População

$AAS_s(2) = \{(12); (13); (21); (23); (31); (32)\}$

↳ espaço amostral

Uma Variável aleatória é

Uma função $X : S \to \mathbb{R}$

$f_1 : AAS_s(2) \to$ 11

$f_1(12) = 1$; $f_1(23) = 0$

$f_1(32) = 1$, ...

Uma distribuição de probabilidades

$P(U) = 1$; $P(\phi) = 0$

$P(A \cup B) \leq P(A) + P(B)$

$P(\sum X_i) = 1$

$(1\,2) \longmapsto \frac{1}{6}$

$(1\,3) \longmapsto 1/6$

$(2\,1) \longmapsto 1/6$

$(2\,3) \longmapsto 1/6$

$(3\,1) \longmapsto 1/6$

$(3\,2) \longmapsto 1/6$

$P(F_1 = 1)$

$$E[X] = \sum x_i P(X = x_i)$$

$$E[f_1] =$$

Exemplos da Aula

População

Exemplo 2.1

Considere a população formada por três domicílios $\mathcal{U}$ e que estão sendo observadas as seguintes variáveis: nome (do chefe), sexo, idade, fumante ou não, renda bruta familiar e número de trabalhadores. Os dados estão na planilha "aula-01-exemplo".

	Nome do Chefe	Sexo	Idade	Fumante	Renda	Nº de Trab
1	Ada	0	20	0	12	1
2	Beto	1	30	1	30	3
3	Ema	0	40	1	18	2

População

Vamos usar as notações

Unidade	i
Nome do Chefe	A_i
Sexo	X_i
Idade	Y_i
Fumante	G_i
Renda Bruta Familiar	F_i
Número de Trabalhadores	T_i

Amostra

Exemplo 2.2

Seja $\mathcal{U} = \{1, 2, 3\}$ (vide planilha "aula-01-exemplo"). Os vetores

$$
\begin{aligned}
\boldsymbol{s}_1 &= (1, 2) \\
\boldsymbol{s}_2 &= (2, 1) \\
\boldsymbol{s}_3 &= (1, 1, 3) \\
\boldsymbol{s}_4 &= (3) \\
\boldsymbol{s}_5 &= (2, 2, 1, 3, 2)
\end{aligned}
$$

são exemplos de amostras ordenadas de $\mathcal{U}$.

Definição 2.3

Uma função $P(\boldsymbol{s})$ definida em $\mathcal{S}(\mathcal{U})$, satisfazendo

$$P(\boldsymbol{s}) \geq 0, \text{ para quaisquer } \boldsymbol{s} \in \mathcal{S}(\mathcal{U})$$

e tal que

$$\sum_{\boldsymbol{s} \in \mathcal{S}(\mathcal{U})} P(s) = 1,$$

é chamado um **planejamento amostral ordenado**.

Planejamento Amostral

Exemplo 2.4

Considere $\mathcal{U} = \{1, 2, 3\}$ (vide planilha "aula-01-exemplo").
Considere os seguintes exemplos de planejamentos amostrais:

Planejamento Amostral

Plano A

$$P(11) = P(12) = P(13) = 1/9$$
$$P(21) = P(22) = P(23) = 1/9$$
$$P(31) = P(32) = P(33) = 1/9$$
$$P(\boldsymbol{s}) = 0, \text{ para as demais } \boldsymbol{s} \in \mathcal{S}.$$

Planejamento Amostral

Plano B

$$P(12) = P(13) = P(21) = P(23) = P(31) = P(32) = 1/6$$

$P(\boldsymbol{s}) = 0$, para as demais $\boldsymbol{s} \in \mathcal{S}$.

$\mathcal{U} = \{1, \dots, 6\}$

Plano T:

i - Sorteie com igual probabilidade elementos de $\mathcal{U}$ até que apareça o elemento 3.

(3); (13); (23);

Plano S

Exemplo 2.5

Considere $\mathcal{U} = \{1, 2, 3\}$ (vide planilha "aula-01-exemplo") e a seguinte regra de sorteio:

i - sorteia-se com igual probabilidade um elemento de $\mathcal{U}$, e anota-se a unidade sorteada;

ii - este elemento é devolvido à população e sorteia-se um segundo elemento do mesmo modo.

Mostre que este é o mesmo plano amostral do Plano A. Este plano é conhecido como **amostragem aleatória simples com reposição** (e será detalhado algumas Aulas adiante).

1- Quais amostras aparecem no plano S?

2- Qual probabilidade está associada à cada uma dessas amostras?

1- Usando as regras i e ii, as amostras do plano S tem tamanho 2.

Uma amostra de S tem formato (xy), com $x, y \in U$; sem nenhuma restrição. Logo

Plano $S = \{(11); (12); (13); (21); (22); (23); (31); (32); (33)\}$ = Plano A.

2- Como x e y são sorteados com mesma probabilidade e o sorteio de y é independente do sorteio de x, $P(xy) = 1/9$ (a mesma do plano A1. Logo os planos são os mesmos.

Definição 2.6

Qualquer característica numérica dos dados correspondentes a amostra $\boldsymbol{s}$ é chamada estatística, ou seja, qualquer função $h(\boldsymbol{d}_s)$ que relaciona as observações da amostra $\boldsymbol{s}$.

Exemplo 2.7

Agora considere os dados na planilha "aula-02-exemplo"e a amostra $\boldsymbol{s} = (12)$. Desse modo, tem-se para o vetor $\begin{pmatrix} F_i \\ T_i \end{pmatrix}$ a seguinte matriz de dados da amostra:

(12)

$$\begin{pmatrix} 12 & 30 \\ 1 & 3 \end{pmatrix}.$$

$$\begin{pmatrix} 12 & 1 \\ 30 & 3 \end{pmatrix}$$

	Nome do Chefe	Sexo	Idade	Fumante	Renda	Nº de Trab
					F	T
1	Ada	0	20	0	12	1
2	Beto	1	30	1	30	3
3	Ema	0	40	1	14	2

As médias

$$\overline{f} = \frac{12+30}{2} = 21 \text{ e } \overline{t} = \frac{1+3}{2} = 2,$$

ou a razão

$$r = \frac{12+30}{1+3} = 10,5$$

são exemplos de estatísticas calculadas na amostra $\boldsymbol{s} = (12)$.

$$r\left(\frac{F}{T}\right) \qquad r\left(\frac{T}{F}\right)$$

$$\bar{F}(12) = \frac{f(1) + f(2)}{2} = \frac{12 + 30}{2} = 21$$

$$\bar{t}(12) = \frac{t(1) + t(2)}{2} = \frac{1 + 3}{2} = 2$$

$$r(12) = \frac{f(1) + f(2)}{t(1) + t(2)} = \frac{12 + 30}{1 + 3} = \frac{42}{4} \quad 10.5$$

Escolhido um plano amostral A, tem-se associado o par $(\mathcal{S}_A, P_A)$ dos respectivos pontos amostrais e suas probabilidades.

Estatísticas e Distribuições Amostrais

Fixada agora uma estatística $h(\boldsymbol{d_s})$, quando $\boldsymbol{s}$ percorre $\mathcal{S}_A$, ter-se-á associado uma variável aleatória $H(\boldsymbol{d_s})$ associada ao par $(\mathcal{S}_A, P_A)$.

Considere também a notação

$$p_h = P_A(\boldsymbol{s} \in \mathcal{S}_A;\, H(\boldsymbol{d_s}) = h),$$

que denota a probabilidade sobre o conjunto de todas as amostras $\boldsymbol{s}$ tais que $H(\boldsymbol{d_s}) = h$.

Conhecendo-se todos os valores de h e as suas respectivas probabilidades, tem-se bem identificada a (distribuição da) variável aleatória H.

Definição 2.8

A distribuição amostral de uma estatística $h(\boldsymbol{d_s})$ segundo um plano amostral A, é a distribuição de probabilidades de $H(\boldsymbol{d_s})$, definida sobre $\mathcal{S}_A$, com função de probabilidade dada por

$$p_h = P_A(\boldsymbol{s} \in \mathcal{S}_A;\, H(\boldsymbol{d_s}) = h) = P(h).$$

Exemplo 2.9

Para os dados na planilha "aula-02-exemplo" com dados amostrais

$$\boldsymbol{D} = \begin{pmatrix} F_i \\ T_i \end{pmatrix} = \begin{pmatrix} 12 & 30 & 18 \\ 1 & 3 & 2 \end{pmatrix}, \; i \in \mathcal{U},$$

considere a estatística $r = h(\boldsymbol{d_s})$ como sendo a razão entre o total da renda familiar e o número de trabalhadores na amostra. Considere também os planos amostrais A e B. Temos as seguintes distribuições amostrais:

Definição 2.10

Um estimador é dito **não-viciado** segundo um plano amostral A se

$$E_A[\hat{\theta}] = \theta.$$

Definição 2.11

O **viés** de um estimador $\hat{\theta}(\boldsymbol{d_s})$ segundo um plano amostral A, é dado por

$$B_A[\hat{\theta}] = E_A[\hat{\theta} - \theta] = E_A[\hat{\theta}] - \theta;$$

e o **erro quadrático médio** por

$$\text{EQM}_A[\theta] = E_A[\hat{\theta} - \theta]^2.$$

Estimadores e Propriedades

Com essas definições verifica-se que

$$\text{EQM}_A[\theta] = \text{Var}_A[\theta] + B_A^2[\hat{\theta}].$$

Estimadores e Propriedades

Observe que para uma amostra particular $\boldsymbol{s}$, a diferença $\hat{\theta}(\boldsymbol{s}) - \theta$ mostra o desvio entre o valor estimado e o valor que se desejaria conhecer, ou seja, o erro cometido pelo uso da amostra e do estimador $\hat{\theta}$ para estimar a quantidade de interesse (parâmetro) θ.

Estimadores e Propriedades

Esse desvio é usualmente conhecido por **erro amostral**. Para dada amostra, o erro amostral só pode ser calculado, na situação improvável de θ ser conhecido.

Estimadores e Propriedades

Por isso, a estratégia da avaliação da amostragem não é julgar o resultado particular de uma amostra, mas do plano amostral. Em outras palavras, queremos avaliar as propriedades do estimador sob a ótica de um plano amostral A.

Exemplo 2.12

Usando os dados da planilha "aula-02-exemplo" (com as notações adotadas até então) temos

$$E_{AASc}(r) \cong 10.13 \text{ e } \mathrm{Var}_{AASc}(r) \cong 0.6289.$$

Suponha que o parâmetro de interesse seja a renda média por trabalhador, R, ou seja,

$$R = \frac{12 + 30 + 18}{1 + 3 + 2} = \frac{60}{6} = 10.$$

Estimadores e Propriedades

Observa-se então que r é um estimador viesado para R, pois

$$E_{AASc}(r) \cong 10.13 \neq 10 = R.$$

O vício é dado por

$$B_{AASc}(r) \cong 10.13 - 10 = 0.13,$$

de modo que

$$\text{EMQ}_{AASc}(r) \cong 0.6289 + 0.13^2 = 0.6458.$$

Exemplo 2.13

Com os mesmos dados do Exemplo anterior, suponha agora que o parâmetro de interesse seja a renda média familiar $\mu_F = 20$. Observe que

$$E_{AASc}(\overline{f}) = 20 \text{ e } \mathrm{Var}_{AASc}(\overline{f}) = 28.$$

Isso implica que $\overline{f}$ não é viciado para μ_F, ou seja, $B_{AASc}(\overline{f}) = 0$, de modo que

$$\mathrm{EMQ}_{AASc}(\overline{f}) = \mathrm{Var}_{AASc}(\overline{f}) = 28.$$

Amostragem Aleatória Simples Com Reposição (AASc)

Exemplo 2.14

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 $\boldsymbol{D} = (12, 30, 18)$. Vamos verificar como se comportam $\overline{y}$ e s^2 com relação as funções paramétricas μ e σ^2 de $\boldsymbol{D}$ para o plano amostral AASs com $n = 2$.

U	F	T
1	12	1
2	30	3
3	18	2

$$\begin{pmatrix} 12 & 1 \\ 30 & 3 \\ 18 & 2 \end{pmatrix}$$

$$\begin{pmatrix} 12 & 30 & 18 \\ 1 & 3 & 2 \end{pmatrix}$$

$$\mu_F = \frac{12+30+18}{3} = \frac{60}{3} = 20$$

$$\mu_T = \frac{1+3+2}{1+2+3} = 1$$

$$R = \frac{12+30+18}{1+3+2} = \frac{60}{6} = 10$$

$$AAS_5(2) = \{(12);(13);(21);(23);(31);(32)\}$$

S	$\bar{F}$	$\bar{r}$
(12)	21	10.5
(13)	15	10
(21)	21	10.5
(23)	24	9.6
(31)	15	10
(32)	24	9.6

Dist. Amostral. F

h	15	21	24
$P(F=h)$	$1/3$	$1/3$	$1/3$

Dist. Amost. r

h	9.6	10	10.5
$P(r=h)$	$1/3$	$1/3$	$1/3$

Dist. Amostral. $\bar{F}$

h	15	21	24
$P(\bar{F}=h)$	1/3	1/3	1/3

$$E[\bar{F}] = \sum h_i \, P(\bar{F}=h_i)$$

$$= 15 \cdot \frac{1}{3} + 21 \cdot \frac{1}{3} + 24 \cdot \frac{1}{3}$$

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

$\Rightarrow$ $\bar{F}$ é não-viesado para $\mu(F)$.

Dist. Amost. r

h	9.6	10	10.5
$P(r=h)$	$1/3$	$1/3$	$1/3$

$$E[r] = \sum h_i \, P(r=h_i)$$

$$= 9.6 \cdot \frac{1}{3} + 10 \cdot \frac{1}{3} + 10.5 \cdot \frac{1}{3}$$

$$= \frac{9.6+10+10.5}{3} = \frac{31}{3} \cong 10.33$$

$10.33 \neq R \Rightarrow r$ é viesada para R

Determinação do Tamanho da Amostra

Exemplo 2.15

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 σ^2 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.

$\rightarrow z_\gamma = 1.96$

$$n \geq \frac{\sigma^2}{(\varepsilon / z_{\alpha})^2} = \frac{463.32}{\left(4 / 1.96\right)^2}$$

$$= \frac{463.32}{(2.04)^2} = \frac{463.32}{4.17} \simeq 111.07$$

$$n = 112$$

$s^2 = 0.7 \times 0.3$

Exemplo 2.16

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 $0,01$?

$\sigma^2 \cong \frac{1}{4}$

$$P\left(\frac{|x - \bar{\mu}|}{s^2/\sqrt{n}} \leq \frac{x_\gamma}{s^2/\sqrt{n}} \right) = \gamma$$

$$IC = (0.3 - 0.13 ; 0.3 + 0.13)$$

Exemplo 2.17

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 $n = 2$.

Normalidade e Intervalo de Confiança

Para esta amostra tem-se que uma estimativa de μ é dada por $\overline{y} = 1,296$. Observa-se também que $s^2 = 2,397$. Usando a aproximação normal, vamos construir um intervalo de confiança para μ com 95% de confiança.

Determinação do Tamanho da Amostra

Para esta amostra tem-se que uma estimativa de μ é dada por $\overline{y} = 1,296$. Observa-se também que $s^2 = 2,397$. Vamos encontrar n tal que $B = 0,05$ e $\gamma = 0,05$.

Exemplo 2.18

Considere a população dos operários faltosos dos exemplos anteriores. Suponha que até 3 faltas (3 dias) em 6 meses seja considerada aceitável. Vamos:

1. construir um intervalo de confiança para P com nível de confiança de 95%;
2. calcular o tamanho da amostra com precisão $B = 0,01$ e nível de confiança de 95%.

Exercícios da Lista

Exemplo 3.1

Considere os dados na planilha "aula-02-exemplo" com dados amostrais

$$\boldsymbol{D} = \begin{pmatrix} F_i \\ T_i \end{pmatrix} = \begin{pmatrix} 12 & 30 & 18 \\ 1 & 3 & 2 \end{pmatrix}, \ i \in \mathcal{U}$$

e plano amostral AASc. Agora considere as estatísticas r e $\overline{f}$. Calcule

$$E_{AASc}[r], E_{AASc}[\overline{f}], \mathrm{Var}_{AASc}[r],$$
$$\mathrm{Var}_{AASc}[\overline{f}], \mathrm{Cov}_{AASc}[r, \overline{f}] \text{ e } \mathrm{Corr}_{AASc}[r, \overline{f}].$$

Exercícios da Lista

Exemplo 3.2

Considere os dados na planilha "aula-02-exemplo" com plano amostral AASc. Verifique que r é um estimador viesado para R.

Exercícios da Lista

Exemplo 3.3

Considere os dados na planilha "aula-02-exemplo" com plano amostral AASc. Verifique que $\overline{f}$ é um estimador não-viesado para μ_F.

Exemplo 3.4

Um plano AASc com $n = 30$ foi adotado em uma área da cidade contendo 14848 residências. O número de pessoas por residência na amostra observada foi

$$\boldsymbol{d} = (5, 6, 3, 3, 2, 3, 3, 3, 4, 4, 3, 2, 7, 4, 3, 5, 4, 4, 3, 3, 4, 3, 3, 1, 2, 4, 3, 4, 2, 4).$$

$\bar{x} = 3,47$; $s^2 = 1,5$ $s = 1,22$

a - Encontre uma estimativa do número médio de pessoas por residência na população e uma estimativa para a variância da estimativa obtida.

b - Encontre um intervalo de confiança para μ.

c - Suponha que seja de interesse uma estimativa duas vezes mais precisa que a obtida com a amostra acima. Qual o tamanho da amostra necessário para tal precisão? 80

$$\frac{X_\gamma}{s^2/\sqrt{n}} = z_\gamma$$

$$x_\gamma = z_\gamma \cdot \frac{s^2}{\sqrt{n}}$$

b - $IC = (3{,}47 - 0{,}54 ; 3{,}47 + 0{,}54)$

$= (2{,}93 ; 4{,}01)$

Comentários Finais

Comentários Finais

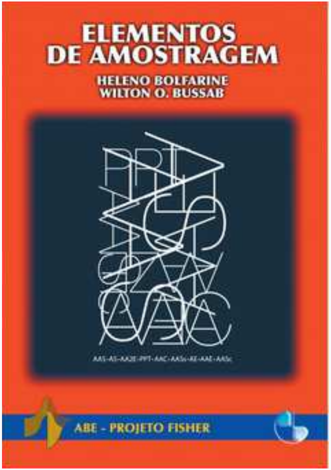
Na aula de hoje nós fizemos um grande resumo da matéria.

Comentários Finais

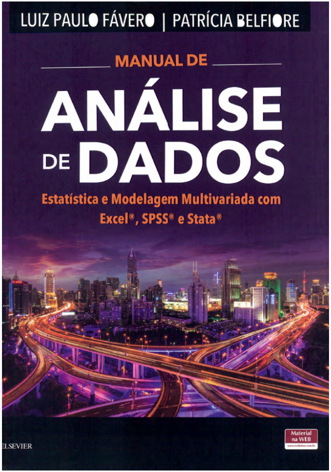
Nas próximas aulas nós vamos começar o estudo da amostragem estratificada.

Referências

Referências



Referências



Bons Estudos!