Técnicas de Amostragem - Aula 05

Amostragem Aleatória Simples Sem Reposição (AASs) II

Kaique Matias de Andrade Roberto

Ciências Atuariais

HECSA - Escola de Negócios

FIAM-FAAM-FMU

Conteúdo

Recomendações para o Trabalho N1

Recomendações para o trabalho N1

O Trabalho devera ser entregue no AvA até o dia 22/10.

Recomendações para o trabalho N1

Para trabalho N1 será cobrado os conteúdos cobertos nas Aulas 00-05.

Recomendações para o trabalho N1

EXERCÍCIOS N1

Dado que os exercícios foram resolvidos em sua maioria durante as aulas, **para um gabarito completo, consulte o material das aulas**.

Recomendações para o trabalho N1

EXERCÍCIOS N1

Segue a lista dos exercícios:

- Aula-00: 0.8, 0.9 e 0.10;
- Aula-01: 1.1, 1.5, 1.7 e 1.10;
- Aula-02: 2.1-2.7;
- Aula-03: 3.5-3.9;
- Aula-04: 4.1-4.3;
- Aula-05: 5.1-5.5.

Recomendações para o trabalho N1

kaique.roberto@fmu.br

EXERCÍCIOS N1

Lembrando que os exercícios podem ser resolvidos em grupos de **até 4 participantes** (para exceções, me procurem o quanto antes), e vocês precisam entregar **apenas uma resolução por grupo**.

Recomendações para o trabalho N1

EXERCÍCIOS N1

O arquivo com as resoluções **será obrigatoriamente do tipo .pdf**. Há várias opções de conversão para arquivos .pdf e caso vocês tenham alguma dificuldade com isso, **me avisem o quanto antes**. **Qualquer extensão diferente de .pdf NÃO será aceita**.

Teórica		Prática	
01	Escolhe 3 questões	08	Escolhe 3 quetões
⋮		⋮	
07		14	

Cada questão vale 2 pontos

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.

- Intervalo de confiança
- Tamanho da amostra

Amostragem Aleatória Simples Sem Reposição (AASs)

- Principais estimadores ($\bar{Y}$, T, s^2)
- Comparação entre os planos

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.

Amostragem Aleatória Simples Sem Reposição (AASs)

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, 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

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

onde $f = n/N$ é a fração amostral.

Teorema 3.4

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

Normalidade e Intervalo de Confiança

Todos os resultados apresentados para o caso com reposição tem o seu equivalente para a AASs, mudando apenas a expressão correspondente à variância amostral. Assim, para a AASs temos os seguintes resultados:

Normalidade e Intervalo de Confiança

$$\frac{\overline{y} - \mu}{\sqrt{(1 - f)s^2/n}} \sim N(0, 1)$$

onde $f = n/N$ é a fração amostral.

Normalidade e Intervalo de Confiança

$$\frac{T - \tau}{N\sqrt{(1-f)s^2/n}} \sim N(0,1)$$

onde $f = n/N$ é a fração amostral.

Normalidade e Intervalo de Confiança

$$P\left(\frac{|\overline{y}-\mu|}{\sqrt{(1-f)s^2/n}} \leq z_\alpha\right) \sim N(0,1),$$

Normalidade e Intervalo de Confiança

resultando no intervalo de confiança μ,

$$\left(\overline{y} - z_\alpha\sqrt{(1-f)\frac{s^2}{n}}; \overline{y} + z_\alpha\sqrt{(1-f)\frac{s^2}{n}}\right),$$

onde $f = n/N$ é a fração amostral.

Normalidade e Intervalo de Confiança

Um intervalo de confiança para τ com coeficiente de confiança aproximadamente $1 - \alpha$ pode ser construído de maneira análoga ao intervalo construído para μ.

Precisão

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

nível de confiança

$$\alpha = 1 - \gamma$$

↳ nível de significância

$N = 36\,000$ $\quad n = 1000$

Exemplo 4.1

Uma pesquisa amostral foi conduzida com o objetivo de se estudar o índice de ausência ao trabalho em um determinado tipo de indústria. Uma AAS sem reposição de mil operários de um total de 36 mil é observada com relação ao número de faltas não justificadas em um período de 6 meses. Os resultados obtidos foram:

$$F = \frac{1000}{36000} = \frac{1}{36}$$

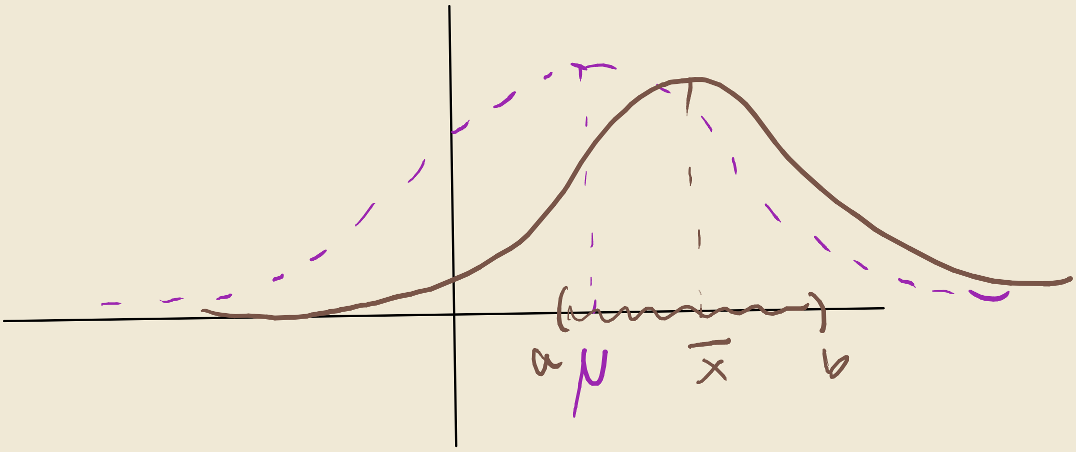
Normalidade e Intervalo de Confiança

Faltas:	0	1	2	3	4	5	6	7	8
Trabalhadores:	451	162	187	112	49	21	5	11	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.

$$P(|\bar{X}-\mu| \leq X_\gamma) = 0.95$$



$\mu \in (a, b)$ com γ de confiança

$$P\left(\boxed{\frac{|\bar{X}-\mu|}{\sqrt{\frac{(1-f)s^2}{n}}}} \leq \frac{X_\gamma}{\sqrt{\frac{(1-f)s^2}{n}}}\right) = 0.95$$

$$P\left(|Z| \leq \underbrace{\boxed{\frac{X_\gamma}{\sqrt{\frac{(1-f)s^2}{n}}}}}_{\to z_\gamma}\right) = 0.95$$

$$P(|Z| \leq z_\gamma) = 0.95$$

$$\boxed{P(-z_\gamma \leq Z \leq z_\gamma) = 0.95}$$

$$z_\gamma = 1.96 \Rightarrow$$

$$\frac{x_\gamma}{\sqrt{\frac{(1-f)s^2}{n}}} = 1.96 \Rightarrow$$

$$x_\gamma = 1.96 \cdot \sqrt{\frac{(1-f)s^2}{n}}$$

$$= 1.96 \sqrt{\frac{\left(1 - \frac{1}{36}\right) 2.397}{1000}} \cong$$

$$\mu \in (1296 - \quad ; 1296 + \quad)$$

Determinação do Tamanho da Amostra

Determinação do Tamanho da Amostra

Pode-se mostrar que o tamanho da amostra para que

$$P(|\overline{y} - \mu| \leq B) = \gamma$$

é dada por

$$n = \frac{1}{D/s^2 + 1/N}, \text{ onde } D = B^2/z_\gamma^2.$$

Determinação do Tamanho da Amostra

Exemplo 5.1

Considere a população dos operários faltosos do exemplo anterior.

Determinação do Tamanho da Amostra

Faltas:	0	1	2	3	4	5	6	7	8
Trabalhadores:	451	162	187	112	49	21	5	11	2

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 $\gamma = 0.05$.

$$P\left(|\bar{x}-\mu| \leq 0.05\right) = 0.95$$

$$P\left(\frac{|\bar{x}-\mu|}{\sqrt{\frac{(1-f)s^2}{n}}} \leq \frac{0.05}{\sqrt{\frac{(1-f)s^2}{n}}}\right) = 0.95$$

$$P\left(|z| \leq \frac{0.05}{\sqrt{\frac{(1-f)s^2}{n}}}\right) = 0.95$$

$$\frac{0.05}{\sqrt{\frac{(1-f)s^2}{n}}} = 1.96 \Rightarrow$$

$$\sqrt{\frac{(1-f)s^2}{n}} = 1.96 \cdot 0.05 \Rightarrow$$

$$\frac{(1-f)s^2}{n} = 1.96^2 \cdot 0.05^2$$

$$n = \frac{(1-f)\,s^2}{1.96^2 \cdot 0.05^2}$$

$$= \frac{\left(1-\frac{1}{36}\right) 2.397}{1.96^2 \cdot 0.05^2} = 297$$

Estimação da Proporção

Estimação da Proporção

De maneira geral, em muitas situações, existe interesse em estudar a proporção de elementos em certa população que possuem determinada característica, como ser ou não um item defeituoso, ser ou não eleitor de um determinado partido, etc.

Estimação da Proporção

Nesta situação, a cada elemento da população está associada a variável aleatória Y_i da seguinte maneira:

$$Y_i = \begin{cases} 1 \text{ se o indivíduo for portador da característica} \\ 0 \text{ se o indivíduo não for portador da característica} \end{cases}$$

Estimação da Proporção

Daí

$$P = \mu = \frac{1}{N}\sum_{i=1}^{N} Y_i$$

é a proporção de unidades da população que possuem a característica de interesse, e podemos escrever

$$S^2 = \frac{N}{N-1}P(1-P).$$

Estimação da Proporção

A seguir segue um resumo dos resultados teóricos obtidos para a estimação da proporção:

Teorema 6.1

Um estimador não-viciado de P baseado na AASs é dado por:

$$p = \hat{P} = \overline{y} = \frac{m}{n},$$

com

$$\text{Var}(\hat{P}) = \frac{N-n}{N-1}\frac{PQ}{n},$$

sendo $Q = 1 - P$. Além disso, um estimador não-viciado de S^2 é

$$s^2 = \frac{n}{n-1}\hat{P}\hat{Q}.$$

Consequentemente, um estimador não viciado de $\text{Var}(\hat{P})$ é dado por

$$\text{var}(p) = (1-f)\frac{\hat{P}\hat{Q}}{n-1}.$$

Utilizando-se a aproximação normal, um intervalo de confiança aproximado para P é dado por

$$\left(\hat{P} - z_{\alpha}\sqrt{(1-f)\frac{\hat{P}\hat{Q}}{n-1}}; \hat{P} + z_{\alpha}\sqrt{(1-f)\frac{\hat{P}\hat{Q}}{n-1}}\right).$$

Notando-se que o produto PQ (e portanto $\hat{P}\hat{Q}$) é sempre menor que $1/4$, um intervalo de confiança conservador para P é dado por

$$\left(\hat{P} - z_\alpha\sqrt{\frac{1-f}{4(n-1)}}; \hat{P} + z_\alpha\sqrt{\frac{1-f}{4(n-1)}}\right).$$

Estimação da Proporção

Como no caso da Média amostral, pode-se considerar o tamanho da amostra n de tal forma que

$$P(|\hat{P} - P| \leq \varepsilon) \cong \gamma.$$

Estimação da Proporção

Pode-se mostrar que o valor de n é dado por

$$n = \frac{N}{\frac{(N-1)D}{PQ} + 1},$$

e para usarmos esta fórmula, é necessário um valor estimado para P.

Estimação da Proporção

Uma forma alternativa, que produz um valor conservador para n consiste em utilizar o fato de que $PQ \leq 1/4$. Neste caso, tem-se

$$n = \frac{N}{4(N-1)D+1}.$$

Exemplo 6.2

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%.

Estimação da Proporção

$$\hat{p} = \frac{451 + 162 + 187 + 112}{1000} = 0.912$$

Faltas:	0	1	2	3	4	5	6	7	8
Trabalhadores:	451	162	187	112	49	21	5	11	2

$$\hat{q} = 1 - \hat{p} = 0.088$$

$$S^2 = \frac{1}{4}$$

$$P\left(|\hat{p} - p| \leq X_\gamma\right) = 0.95$$

$$P\left(\frac{|\hat{p} - p|}{\sqrt{\frac{(1-f)\,\hat{p}\hat{q}}{n-1}}} \leq \frac{X_\gamma}{\sqrt{\frac{(1-f)\,\hat{p}\hat{q}}{n-1}}}\right) = 0.95$$

$$\frac{X_\gamma}{\sqrt{\frac{(1-f)\,\hat{p}\hat{q}}{n-1}}} = 1.96$$

AASs versus AASc

AASs versus AASc

Quando se tem dois planos amostrais é importante saber qual é "melhor". Para isso, é preciso fixar o critério pelo qual o plano será julgado.

AASs versus AASc

Dentre os critérios mais adotados estão o Erro Quadrático Médio ou a variância quando os estimadores não são viesados.

AASs versus AASc

Além disso, existe um conceito importante, o **Efeito do Planejamento, EPA,** que compara a variância de um plano qualquer com relação a um plano que é considerado padrão.

AASs versus AASc

A estatística $\overline{y}$ é em ambos os planos (AASs e AASc) um estimador não viesado. Assim

$$\text{EPA} = \frac{\text{Var}_{AASs}[\overline{y}]}{\text{Var}_{AASc}[\overline{y}]} = \frac{(1-f)S^2/n}{\sigma^2/n} = \frac{N-n}{N-1}.$$

AASs versus AASc

Quando o EPA > 1 tem-se que o plano do numerador é menos eficiente que o padrão. Quando EPA < 1 a situação é inversa.

AASs versus AASc

$N - n \leq N - 1$

No nosso caso,

$$\text{EPA} = \frac{\text{Var}_{AASs}[\overline{y}]}{\text{Var}_{AASc}[\overline{y}]} = \frac{N - n}{N - 1} \leq 1$$

com a igualdade acontecendo apenas quando $n = 1$.

AASs versus AASc

Ou seja, o plano AASs é "melhor" do que o plano AASc. Esse resultado confirma a intuição de que amostras sem reposição são "melhores" do que aquelas com elementos repetidos.

Comentários Finais

Comentários Finais

Em resumo, na aula de hoje nós completamos a lista de propriedades da AASs e com isso, finalizamos o desenvolvimento dos aspectos básicos da teoria envolvendo AAS (com ou sem reposição).

Comentários Finais

Nas próximas aulas nós vamos focar em números aleatórios.

Comentários Finais

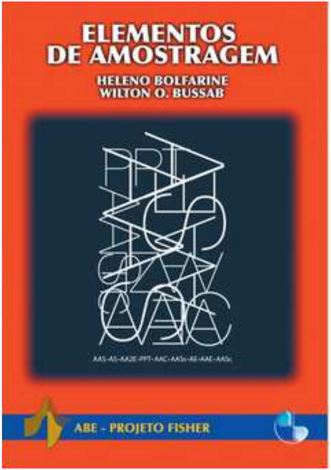
REFORÇO: EXERCÍCIOS N1

Segue a lista dos exercícios:

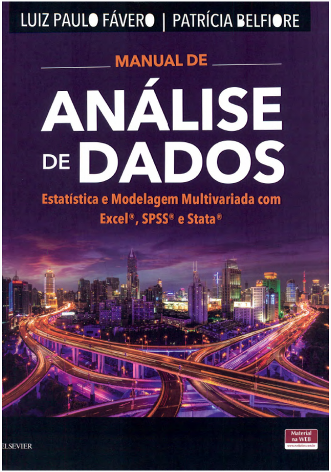
- Aula-00: 0.8, 0.9 e 0.10;
- Aula-01: 1.1, 1.5, 1.7 e 1.10;
- Aula-02: 2.1-2.7;
- Aula-03: 3.5-3.9;
- Aula-04: 4.1-4.3;
- Aula-05: 5.1-5.5.

Referências

Referências



Referências



Bons Estudos!

$\boxed{2.4}$ $D = \begin{pmatrix} F_i \\ T_i \end{pmatrix} = \begin{pmatrix} 12 & 30 & 18 \\ 1 & 3 & 2 \end{pmatrix}$

$AASc(2)$ $r = \frac{\Sigma F_i}{\Sigma T_i}$ $P_{AAc(2)}(s) = \frac{1}{9}$

S	11	12	13	21	22	23	31	32	33
$\bar{F}$	12	21	15	21	30	24	15	24	18
$\bar{T}$	1	2	1.5	2	3	2.5	1.5	2.5	2
r	12	10.5	10	10.5	10	9.6	10	9.6	9

h	12	15	18	21	24	30
$P(\bar{f}=h)$	$1/9$	$2/9$	$1/9$	$3/9$	$2/9$	$1/9$

$$E[\bar{f}] = 12\cdot\frac{1}{9} + 15\cdot\frac{2}{9} + \cdots + 30\cdot\frac{1}{9}$$

$$= 20 = \mu$$

h	9	9.6	10	10.5	12
$P(r=h)$	$1/9$	$2/9$	$3/9$	$2/9$	$1/9$

$$E[r] = 10.23$$

$$\mathrm{Cov}[\bar{f}, r] =$$

$$\sum_{s \in AAS_c(z)} \left[\bar{f}(s) - E[\bar{f}] \right] \left[r(s) - E[r] \right] \cdot P(s)$$

$$= \frac{1}{9} \Big[(12-20)(1-10.13) + (23-20)(2-10.3) + \ldots$$

$$+ (18-20)(2-10.13) \Big] = \sigma_{21}$$

$$\mathrm{Correl}[\bar{f}, r] = \frac{\sigma_{21}}{\sqrt{\mathrm{Var}[\bar{f}]\,\mathrm{Var}[\bar{r}]}}$$