ME 430A - Técnicas de Amostragem
Segundo semestre de 2011
Prova II
Data: 17/10/2011

Nome: ________________________________ RA: ______________

Leia atentamente as instruções abaixo:

- **Coloque seu nome completo e RA em todas as folhas que você recebeu, inclusive nesta.**
- **Utilize somente um dos lados de cada folha.**
- **Leia atentamente cada uma das questões.**
- **Enuncie, claramente, todos os resultados que você utilizar.**
- **Justifique, adequadamente, seus desenvolvimentos, sem, no entanto, escrever excessivamente.**
- **O(a) aluno(a) só poderá sair da sala após as 16h30, mesmo que já tenha finalizado a prova. Após a saída do(a) primeiro(a) aluno(a) não será permitido a entrada de nenhum(a) outro(a) aluno(a).**
- **Não é permitdo empréstimo de material.**
- **Não serão dirimidas dúvidas de quaisquer natureza.**
- **Resolva a prova, preferencialmente, à caneta, e procure ser organizado(a). Se fizer à lápis, destaque, à caneta, sua resposta.**
- **O(a) aluno(a) deverá portar sua carteira de estudante e apresentá-la, quando for solicitada sua assinatura.**
- **Contestações a respeito da nota, só serão consideradas se estiverem por escrito.**
- **A nota do aluno(a) será $\frac{NP}{NT} \times 10$, em que NP é o número de pontos obtidos na prova e NT é o numero total de pontos da prova.**
- **Os resultados numéricos devem ser apresentados com duas casas decimais, apenas.**
- **A prova terá duração de 120 minutos, das 16h às 18h, improrrogáveis.**

Faça uma excelente Prova!!

Questões

1. Considere uma população dividida em H estratos, com N_h elementos em cada uma deles, com $N = \sum_{h=1}^{H} N_h$ o tamanho da população e $W_h = \frac{N_h}{N}$. Suponha que, de cada um dos estratos, é selecionada uma amostra, de tamanho $n_h, h = 1, ..., H$, através do esquema AASs (sem reposição) de modo independente entre os estratos. Seja ainda $n = \sum_{h=1}^{H} n_h$ o tamanho da amostra (total). Responda os itens:

   a) Seja $\widehat{\mu}_{es} = \displaystyle\sum_{h=1}^{H} W_h \widehat{\mu}_h$, um estimador para μ, em que $\widehat{\mu}_h = \dfrac{1}{n_h} \displaystyle\sum_{i \in \boldsymbol{s}_h} Y_{hi}$. Prove que esse estimador é não viciado para μ e que $\mathcal{V}_{AE_2}(\widehat{\mu}_{es}) = \displaystyle\sum_{h=1}^{H} W_h^2 \left(1 - \frac{n_h}{N_h}\right) \frac{s_h^2}{n_h}$. (30 pontos)

   b) Considere uma população divida em 3 estratos, com as seguintes características, para uma amostra estratificada de tamanho 50, selecionada segundo um AASs (sem reposição) dentro de cada estrato, tal como mostrada na tabela abaixo:

N_h	W_h	s_h^2	n_h	$\widetilde{\mu}_h$	$W_h\widetilde{\mu}_h$	f_h	$W_h^2(1-f_h)$	$W_h^2(1-f_h)\frac{s_h^2}{n_h}$
300	0,30	10,00	15,00	2,00	0,60	0,05	0,09	0,06
500	0,50	20,00	30,00	3,00	1,50	0,06	0,24	0,16
200	0,20	5,00	5,00	4,00	0,80	0,03	0,04	0,04

   Obtenha a estimativa, o erro-padrão e um IC assintótico com c.c. aproximado de $\gamma = 0,95$, para μ, utilizando o estimador descrito no item a) ($\widehat{\mu}_{es}$).(50 pontos)

   c) Para os dados do item b), obtenha a estimativa, o erro-padrão e o IC assintótico, equivalente ao item b), utilizando $\widehat{\mu} = \frac{1}{n}\sum_{i \in \boldsymbol{s}} Y_i$. Compare a estimativa obtida com aquela do item b) e diga, justificando adequadamente, qual das duas você escolheria para fazer inferências sobre μ. Justifique, adequadamente, como obter a estimativa com $\widehat{\mu}$ bem como para o cálculo do erro-padrão associado.(50 pontos)

   d) Suponha, para a população descrita no item a): alocação uniforme entre os estratos, $N_h = \frac{N}{H}$ e que $s_h^2 = \alpha_i s^2, \alpha_h \in (0,1), \forall h$, $h = 1, ..., H$ e que $\sum_{h=1}^{H} \alpha_i < H$. Compare o estimador do item a) com o estimador usual $\widehat{\mu}$ sob AASs (sem reposição), em relação às suas variâncias. Qual dos dois estimadores você usuaria? Justifique, adequadamente, sua resposta.(70 pontos)

2. Considere uma população com 3 elementos (domicílios), $\mathcal{U} = \{1, 2, 3\}$, dos quais temos as informações sobre a renda do domicílio (R) e o número de trabalhadores do domicílio (T). A tabela abaixo descreve um plano amostral, para n = 2, associado à população em questão, bem como a distribuição amostral de dois estimadores $\overline{R}$ e $\overline{R}_R$ para μ_r (média populacional da renda). $\overline{R}$ é a média amostral de R, $\overline{R}_R = \mu_t \frac{\overline{R}}{\overline{T}}$, em que μ_t é a média populacional de T e $\overline{T}$ é a média amostral de T.

$\boldsymbol{s}$	12	13	23
$P(\boldsymbol{s})$	0,33	0,33	0,33
$\overline{R}$	21,00	15,00	24,00
$\overline{R}_R$	21,00	20,00	19,20

Responda os itens.

a) Calcule $\mathcal{E}(\overline{R}), \mathcal{V}(\overline{R}), \mathcal{EQM}(\overline{R}), \mathcal{E}(\overline{R}_R), \mathcal{V}(\overline{R}_R), \mathcal{EQM}(\overline{R}_R)$, sob o plano amostral em questão, para estimar a renda média populacional dos domicílios μ_r. Considere que $\mu_r = 20,0$ e $\mu_t = 2,0$.(100 pontos)

b) Compare $\overline{R}$ e $\overline{R}_R$ através de seus EQM's. Qual dos dois estimadores você escolheria para estimar μ_r, justifique, adequadamente, sua resposta.(50 pontos)

c) Considere uma população com N elementos, da qual se seleciona uma amostra de tamanho n, segundo um esquema AASs (sem reposição). Seja $\widehat{\mu}_R = \mu_x \widehat{r}$, o estimador razão para média populacional μ. Prove (sem fazer contas em detalhes, ou seja, apenas argumentando) que

$\mathcal{E}_{A_2}(\widehat{\mu}_R) \approx \mu_y$ e $\mathcal{V}_{A_2}(\widehat{\mu}_R) \approx \left(1 - \frac{n}{N}\right) \frac{s_R^2}{n}$, em que $s_R^2 = \frac{1}{n-1} \sum_{i \in \boldsymbol{s}} (y_i - r x_i)^2$.(50 pontos)

3. Considere uma população dividida em 3 estratos, como se segue:

W_h	N_h
0,1	500
0,6	3000
0,3	1500

Responda os itens:

a) Considerando $n = 210$, construa uma tabela com os valores de n_h para as alocações proporcional e uniforme.(30 pontos)

b) Com base em cada uma das alocações, obtenha estimativas para p utilizando o estimador $\widehat{p}_{es} = \sum_{h=1}^{H} W_h \widehat{p}_h$, bem como os erros-padrão e os IC's assintóticos com $\gamma = 0,99$. Qual das alocações você utilizaria para estimar p? Justifique, adequadamente, sua resposta. Para isso, considere que para cada uma das alocações, com os tamanhos de amostra que você obteve, foram obtidos $\widetilde{p}_1 = 0,20$, $\widetilde{p}_2 = 0,30$ e $\widetilde{p}_3 = 0,45$. Sugestão: escreva as contas necessárias na tabela que você construiu no item anterior.(70 pontos)

c) Considerando uma população geral (como a da questão 1 a)), dividida em H estratos e supondo alocação proporcional com AASc (com reposição) em cada estrato, prove que o tamanho da amostra tal que

$$P_{AE_2}(|\widehat{p}_{es} - p| < \delta) = \gamma$$

é dado por:

$$n = \frac{z^2}{\delta^2}\left(\sum_{h=1}^{H} W_h p_h (1 - p_h)\right),$$

em que $P(Z > z) = \frac{1-\gamma}{2} Z \sim N(0,1)$.(70 pontos)

d) Considerando a população em questão (do enunciado): qual o tamanho de amostra, de acordo com o item c), considerando as estimativas para $p_h, h = 1, 2, 3$ obtidas no item b), $\delta = 0,02$ e $\gamma = 0,90$.(30 pontos)

Formulário

Amostragem Aleatória Simples

1. Parâmetros populacionais de interesse: $\mu = \frac{1}{N}\sum_{i=1}^{N} y_i$; $\tau = N\mu = \sum_{i=1}^{N} y_i$; $\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(y_i - \mu)^2$; $s^2 = \frac{1}{N-1}\sum_{i=1}^{N}(y_i - \mu)^2$, $p = \frac{1}{N}\sum_{i=1}^{N} y_i, y_i \in \{0, 1\}$.

2. Estimadores: $\widehat{\mu} = \frac{1}{n}\sum_{i \in \boldsymbol{s}} Y_i$;$\widehat{\tau} = N\widehat{\mu} = \sum_{i \in \boldsymbol{s}} Y_i$; $\widehat{\sigma}^2 = \widehat{s}^2 = \frac{1}{n-1}\sum_{i \in \boldsymbol{s}}(Y_i - \widehat{\mu})^2$, $\widehat{p} = \frac{1}{n}\sum_{i \in \boldsymbol{s}} Y_i, Y_i \in \{0, 1\}$.

3. Variâncias dos estimadores

   (a) AASc : $\mathcal{V}(\widehat{\mu}) = \frac{\sigma^2}{n}$; $\mathcal{V}(\widehat{\tau}) = N^2\frac{\sigma^2}{n}$; $\mathcal{V}(\widehat{p}) = \frac{p(1-p)}{n}$.

   (b) AASs : $\mathcal{V}(\widehat{\mu}) = \left(1 - \frac{n}{N}\right)\frac{s^2}{n}$; $\mathcal{V}(\widehat{\tau}) = \left(1 - \frac{n}{N}\right)N^2\frac{s^2}{n}$; $\mathcal{V}(\widehat{p}) = \left(\frac{N-n}{N-1}\right)\frac{p(1-p)}{n}$.

4. Estimadores não viciados para as variâncias dos estimadores

   (a) AASc : $\widehat{\mathcal{V}}(\widehat{\mu}) = \frac{\widehat{\sigma}^2}{n}$; $\widehat{\mathcal{V}}(\widehat{\tau}) = N^2\frac{\widehat{\sigma}^2}{n}$; $\widehat{\mathcal{V}}(\widehat{p}) = \frac{\widehat{p}(1-\widehat{p})}{n-1}$.

   (b) AASs : $\widehat{\mathcal{V}}(\widehat{\mu}) = \left(1 - \frac{n}{N}\right)\frac{\widehat{s}^2}{n}$; $\widehat{\mathcal{V}}(\widehat{\tau}) = \left(1 - \frac{n}{N}\right)N^2\frac{\widehat{s}^2}{n}$; $\widehat{\mathcal{V}}(\widehat{p}) = \left(\frac{1-n/N}{n-1}\right)\widehat{p}(1 - \widehat{p})$.

5. Parâmetros populacionais de interesse:

   $\mu = \frac{1}{N}\sum_{i=1}^{H} W_h\mu_h$; $\mu_h = \frac{1}{N_h}\sum_{i=1}^{N_h} y_{hi}$; $\tau = N\mu$;
   $\sigma^2 : \frac{1}{N}\sum_{h=1}^{H}\sum_{i=1}^{N_h}(y_{hi}-\mu)^2 = \sum_{h=1}^{H} W_h\sigma_h^2 + \sum_{h=1}^{H} W_h(\mu_h-\mu)^2$; $\sigma_h^2 = \frac{1}{N_h}\sum_{i=1}^{N_h}(y_{hi}-\mu_h)^2$;
   $s^2 = \frac{1}{N-1}\sum_{h=1}^{H}\sum_{i=1}^{N_h}(y_{hi}-\mu)^2 = \sum_{h=1}^{H}\frac{N_h-1}{N-1}s_h^2 + \sum_{h=1}^{H}\frac{N_h}{N-1}(\mu_h-\mu)^2$; $s_h^2 = \frac{1}{N_h-1}\sum_{i=1}^{N_h}(y_{hi}-\mu_h)^2$,
   $p = \frac{1}{N}\sum_{i=1}^{H} W_h p_h; p_h = \dfrac{1}{N_h}\sum_{i=1}^{N_h} y_{hi}, y_{hi} \in \{0,1\}$.

6. Estimadores: $\widehat{\mu}_{es} = \sum_{h=1}^{H} W_h\widehat{\mu}_h; \widehat{\mu}_h = \frac{1}{n_h}\sum_{i\in \boldsymbol{s}_h} Y_{hi}; \widehat{\tau}_{es} = N\widehat{\mu}_{es}; \widehat{\sigma}^2 = \widehat{s}^2 = \sum_{h=1}^{H} W_h\widehat{s}_h^2; \widehat{s}_h^2 = \frac{1}{n_h-1}\sum_{i\in \boldsymbol{s}_h}(Y_{hi}-\widehat{\mu}_h)^2$, $\widehat{p} = \sum_{h=1}^{H} W_h\widehat{p}_h, \widehat{p}_h = \frac{1}{n_h}\sum_{i\in \boldsymbol{s}_h} Y_{hi}, Y_{hi} \in \{0,1\}$

7. Variância dos estimadores: basta utilizar as propriedades da variância da soma de variáveis aleatórias independentes e os resultados sob AAS, por exemplo:

   $\mathcal{V}_{AE_1}(\widehat{\mu})_{es} = \sum_{h=1}^{H} W_h^2\frac{\sigma_h^2}{n_h}$, $\mathcal{V}_{AE_2}(\widehat{\mu}_{es} = \sum_{h=1}^{H} W_h^2\left(1-\frac{n_h}{N_h}\right)\frac{s_h^2}{n_h}$

8. Estimadores não viciados das variâncias dos estimadores: basta substituir as quantidades desconhecidas pelos respectivos estimadores não viciados, por exemplo:

   $\widehat{\mathcal{V}}_{AE_1}(\widehat{\mu})_{es} = \sum_{h=1}^{H} W_h^2\frac{\widehat{\sigma}_h^2}{n_h}$, $\widehat{\mathcal{V}}_{AE_2}(\widehat{\mu})_{es} = \sum_{h=1}^{H} W_h^2\left(1-\frac{n_h}{N_h}\right)\frac{\widehat{s}_h^2}{n_h}$

Estimador razão

9. Estimadores: $\widehat{r} = \frac{\widehat{\mu}_y}{\widehat{\mu}_x}$, $\widehat{\mu}_R = \mu_x\widehat{r}$, $\widehat{\tau}_R = \tau_x\widehat{r}$. Os estimadores sob AE são obtidos de modo semelhante aos estimadores convencionais.

10. Variâncias: são exatamente as mesmas obtidas sob AAS e AE, substituindo σ^2 e s^2, por σ_R^2 e s_R^2, respectivamente, por exemplo:

    $\mathcal{V}(\widehat{\mu}_R) \approx \frac{\sigma_R^2}{n}$, $\mathcal{V}(\widehat{\mu}_R) \approx \left(1-\frac{n}{N}\right)\frac{s_R^2}{n}$, $\sigma_R^2 = \frac{1}{N}\sum_{i=1}^{N}(y_i - rx_i)^2$, $s_R^2 = \frac{1}{N-1}\sum_{i=1}^{N}(y_i - rx_i)^2$.

11. Estimadores para as variâncias: basta substituit r na fórmula por $\widehat{r}$, por exemplo:

    $\widehat{\mathcal{V}}(\widehat{\mu}_R) = \frac{\widehat{\sigma}_R^2}{n}$, $\widehat{\mathcal{V}}(\widehat{\mu}_R) = \left(1-\frac{n}{N}\right)\frac{\widehat{s}_R^2}{n}$, $\widehat{\sigma}_R^2 = \frac{1}{n}\sum_{i\in \boldsymbol{s}}(Y_i - \widehat{r}X_i)^2$, $\widehat{s}_R^2 = \frac{1}{n-1}\sum_{i\in \boldsymbol{s}}(Y_i - \widehat{r}X_i)^2$.