Departamento de Matemática - Universidade de Coimbra
Exame de Recurso de Amostragem e Sondagens – 2007/08

10/7/08 tempo: 2h30m

Atenção: *Indique de forma clara, na página de rosto da folha de prova, qual a lista de números aleatórios que lhe foi entregue. De todas as vezes que tiver necessidade de utilizar essa lista, indique na sua prova qual o número que leu. A lista deve ser consultada da esquerda para a direita e de cima para baixo.*

1. Na população $\mathcal{U} = \{1, 2, 3, 4, 5, 6, 7, 8\}$, com $\mathbf{Y} = (8, 7, 6, 5, 4, 3, 2, 1)$, pretendemos seleccionar uma amostra de tamanho 2, sem repetições e ignorando a ordem das unidades seleccionadas. Pretende-se que as amostras $(1, 2)$, $(1, 4)$, $(2, 3)$, $(3, 6)$, $(4, 7)$ e $(7, 8)$ sejam seleccionadas com probabilidade p, enquanto que as restantes são seleccionadas com probabilidade $2p$.

   (a) Mostre que $p = 0.02$.

   (b) Descreva as probabilidades de inclusão de primeira ordem.

   (c) Suponha que foi seleccionada a amostra (1,6). Indique uma estimativa para o total da população. Indique uma aproximação para a variância associada a esta estimativa.

2. Na população $\mathcal{U} = \{1, 2, 3, 4, 5\}$ executa-se um plano de amostragem de acordo com as seguintes regras, onde $\alpha \in (0, 1)$ é um número fixo:

   - com probabilidade α executa-se um SRS$(3, 2)$ em $\{1, 2, 3\}$;
   - com probabilidade $1 - \alpha$ executa-se um SRS$(3, 2)$ em $\{3, 4, 5\}$.

   (a) Descreva o suporte deste plano de amostragem.

   (b) Calcule as probabilidades de inclusão de primeira e segunda ordens.

   (c) Mostre que este plano de amostragem permite ter a garantia de obter estimativas para a variância do estimador de Horvitz-Thompson que são sempre não negativas.

   (d) O estimador para a variância é não enviesado?

   (e) Admita que $\alpha = 0.43$. Escolha uma amostra de acordo com este plano de amostragem. Admitindo que $\mathbf{Y} = (-1, -2, 10, 1, 3)$, indique uma estimativa para a variância do estimador de Horvtiz-Thompson.

   (f) Utilizando a amostra da alínea anterior, indique uma estimativa para a variância da população.

3. Para determinar o tamanho N de uma população de ratos coloca-se uma armadilha no local onde estes habitam que captura cada animal sem o molestar. Estes animais são marcados de forma não intrusiva e novamente libertados. Marcam-se desta forma 200 ratos. Posteriormente, coloca-se a armadilha noutro local seleccionado ao acaso e capturam-se ratos até que a armadilha contenha 70 animais. Destes, verifica-se que 21 estavam marcados. Admite-se que os ratos se deslocam ao acaso no território por eles habitado, sendo portanto a probabilidade de captura uniforme para cada um deles.

   (a) Justifique que o modelo de plano de amostragem na recolha do 50 ratos é convenientemente descrito por um SRS$(N, 50)$.

   (b) Defina, para cada rato i, a variável de interesse $Y_i = 1$ se este rato estiver marcado, e $Y_i = 0$ caso contrário. Obtenha uma estimativa não enviesada para a proporção de ratos marcados na população.

V.S.F.F.

(c) Qual a variância associada à estimativa obtida na alínea anterior? (**Nota**: Caso não responda a esta questão e necessite deste valor em alíneas seguintes admita que é igual a 0.006.)

(d) Construa uma estimativa para o total da população. Indique um intervalo con confiança 95% para esta estimativa.

(e) Pretende-se colocar a armadilha noutro local e recolher uma nova amostra. Que tamanho mínimo poderá aconselhar para que se obtenha um intervalo com confiança 95% para a proporção de ratos marcados cuja amplitude seja, no máximo, 0.075?

4. Uma região habitacional divide-se em três subconjuntos disjuntos $\mathcal{U}_1$, $\mathcal{U}_2$ e $\mathcal{U}_3$, contendo, respectivamente, 9876, 12345 e 6745 núcleos habitacionais. Uma empresa de serviços telefónicos pretende obter uma estimativa do total de horas de utilização diária de serviço ADSL nesta região. A única informação de que dispõe diz respeito a estimativas anteriores sobre a proporção de núcleos que contrataram serviços deste tipo e que são, respectivamente, 0.12, 0.86 e 0.56. Pretende-se efectuar um plano de amostragem estratificado, já que as proporções de assinatura de serviços ADSL parecem ser indicadoras de comportamentos distintos em cada uma dos conjuntos $\mathcal{U}_1$, $\mathcal{U}_2$ e $\mathcal{U}_3$. Pretende-se recolher uma amostra com tamanho total 250, o que corresponde a cerca de 1% do número de núcleos registados.

   (a) Como aconselha a proceder à estratificação? Proporcional ou óptima?Com que tamanhos de amostragem por estrato? Justifique.

   (b) Admita que o número médio de horas de utilização diária é, para cada um dos estratos, 1.2, 5.6, 14.3, respectivamente, com variâncias 0.03, 1.43 e 5.61. Construa estimativas para o valor médio do número de horas de utilização diária e respectiva variância.

5. Considere um plano de amostragem numa população $\mathcal{U}$ de tamanho N. Denotemos por π_i e $\pi_{i,j}$ as probabilidades de inclusão de primeira e segunda ordens, respectivamente, e defina-se o estimador

$$\widehat{\theta}(s,\mathbf{Y}) = \frac{1}{N}\sum_{i\in s}\frac{Y_i}{\pi_i} + \frac{1}{N}\sum_{i,j\in s,\, i\neq j}\frac{Y_j}{\pi_{i,j}}\,.$$

   Mostre que este estimador é não enviesado para o total da população.

6. Considere uma população com N unidades e dois planos de amostragem: d_1 um ΠPS$(N, n_1, \mathbf{p}_1)$ e d_2 um ΠPS$(N, n_2, \mathbf{p}_2)$, onde n_1, $n_2 > 1$ são números naturais menores que N e $\mathbf{p}_1$, $\mathbf{p}_2$ são vectores de pesos normalizados. Fixe-se $\varepsilon \in (0,1)$. Executa-se o seguinte plano de amostragem: com probabilidade ε procede-se à escolha da amostra segundo d_1, com probabilidade $1-\varepsilon$ utiliza-se d_2.

   (a) O plano de amostragem assim construído é um ΠPS? Em que casos é que este plano é ainda um ΠPS?

   (b) Caracterize as probabilidades de inclusão de primeira ordem para o plano definido.

   (c) Suponhamos que $N = 123$, $n_1 = 8$, $n_2 = 6$ e $\varepsilon = 0.12$. Ao executar o plano d_1 seleccionamos a amostra $s_1 = (46, 52, 53, 64, 71, 72, 85, 100)$, enquanto que ao executar o plano d_2 seleccionamos a amostra $s_2 = (18, 64, 77, 82, 108, 117)$. Utilize a lista de números aleatórios para escrever a expressão da estimativa para o total da população (em função de $\mathbf{Y}$, $\mathbf{p}_1$ e $\mathbf{p}_2$, naturalmente).