Departamento de Matemática - Universidade de Coimbra

Exame de Amostragem e Sondagens – 2007/08

18/6/08 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. As seguintes funções descrevem um algoritmo de escolha aleatória:

$q_1(1) = 0.2$	$q_1(2) = 0.7$	$q_1(3) = 0.1$	
$q_2(1) = 0$	$q_2(2) = 1$	$q_2(3) = 0.2$	
$q_3(1\|2) = 0.3$	$q_3(3\|2) = 0.7$	$q_3(1\|3) = 0.8$	$q_3(2\|3) = 0.2$
$q_2(2,1) = 0$	$q_2(2,3) = 0$	$q_2(3,1) = 0$	$q_2(3,2) = 1$
$q_3(1\|(3,2)) = 1$			

   (a) Utilize a lista de números aleatórios para seleccionar uma amostra.

   (b) O plano de amostragem é de tamanho fixo? Justifique.

   (c) O plano de amostragem é reduzido? Justifique.

2. Na população $\mathcal{U} = \{1, 2, 3, 4\}$ definimos o seguinte plano de amostragem:

S	(1)	(2)	$(3,1)$	$(3,2)$	$(3,4)$	$(4,1)$	$(4,2)$	$(4,3)$
P	0.172	0.120	0.058	0.041	0.154	0.144	0.100	0.211

   Admita que $\mathbf{Y} = (10, 15, -1, 0)$.

   (a) Utilize a lista de números aleatórios para seleccionar uma amostra de acordo com o plano descrito.

   (b) Indique estimativas para o total da população e para a variância associada a essa estimativa. São não enviesadas?

   (c) Estime a variância da população normalizada por $N-1$, $S^2_{\mathbf{Y}}$.

   (d) Supondo que dispomos da informação de que a unidade 4 pouco influi na determinação do total da população resolvemos propor o seguinte estimador:

S	(1)	(2)	$(3,1)$	$(3,2)$	$(3,4)$	$(4,1)$	$(4,2)$	$(4,3)$
ℓ_1	$\frac{Y_1}{0.374}$	$\frac{Y_2}{0.261}$	$\frac{Y_1}{0.374} + \frac{Y_3}{0.464}$	$\frac{Y_2}{0.261} + \frac{Y_3}{0.464}$	$2Y_3 + Y_4$	$\frac{Y_1}{0.374} + \frac{Y_4}{0.609}$	$\frac{Y_2}{0.261} + \frac{Y_4}{0609}$	$3Y_3 + 2Y_4$

   Justifique que ℓ_1 é enviesado para o total da população. Defina um estimador ℓ_2, a partir de ℓ_1 que seja não enviesado.

   (e) Caracterize a versão reduzida do plano de amostragem.

   (f) Construa a versão simetrizada do estimador para o total ℓ_1 definido na alínea (d).

3. Uma população está dividida em dois grupos disjuntos $\mathcal{U}_1$ e $\mathcal{U}_2$ com N_1 e N_2 unidades, respectivamente. Representemos por $N_0 = \max(N_1, N_2)$. Procede-se à construção de uma amostra com duas unidades da seguinte forma:

   - escolhe-se um números ao acaso u_1 no intervalo [0,1];
   - se $u_1 < 0.3$ então seleccionam-se duas unidades em $\mathcal{U}_1$; caso contrário seleccionam-se as duas unidades em $\mathcal{U}_2$;
   - para seleccionar as duas unidades em $\mathcal{U}_i$ repete-se o números de vezes que for necessário os seguintes passos:

- escolher um número inteiro k ao acaso entre 1 e N_0; se $k > N_i$ ignora-se esta escolha e repete-se este passo;
- inclui-se na amostra a unidade k; caso esta seja a segunda unidade a incluir e seja igual à primeira ignora-se o resultado e repete-se a selecção das duas unidades.

(a) Descreva o plano de amostragem definido, isto é, caracterize o suporte e calcule a probabilidade associada à selecção de cada amostra.

(b) Justifique que este plano de amostragem não permite a construção de bons intervalos de confiança para o total da população.

(c) Suponha que $N_1 = 10$ e $N_2 = 13$. Utilize a lista de números aleatórios para proceder à escolha de uma amostra de acordo com o procedimento descrito.

4. Um departamento é responsável por 186 disciplinas sendo que 40 destas têm menos de dez alunos inscritos. No que se segue, estas serão designadas por *disciplinas pequenas* sendo as restantes classificadas como *disciplinas grandes*. Pretende-se avaliar o grau de satisfação dos estudantes com as disciplinas à responsabilidade do departamento.

(a) Escolhem-se, ao acaso, 14 disciplinas para consulta aos estudantes nelas inscritos. Destas, 4 são disciplinas pequenas e as restantes 10 são disciplinas grandes. Nas disciplinas pequenas incluem-se na amostra todos os estudantes inscritos enquanto que nas disciplinas grandes se efectuam amostragens aleatórias simples para definição da amostra. A tabela seguinte descreve os números de estudantes inscritos nas 14 disciplinas seleccionadas, os tamanhos das amostragens efectuadas em cada uma delas e o total de estudantes que se mostraram satisfeitos na amostra referente a cada disciplina.

N_i	7	5	9	8	52	104	23	35	18	150	78	223	19	65
n_i	7	5	9	8	5	10	3	4	2	15	8	23	2	7
$\widehat{T}_i$	7	3	5	6	3	6	1	2	1	6	7	15	0	3

Construa uma estimativa para o total de alunos satisfeitos com o funcionamento das disciplinas.

(b) Indique uma aproximação para a variância associada à estimativa obtida na alínea anterior

(c) A soma do número de estudantes inscritos às 186 disciplinas é 7038. Acha razoável afirmar que a proporção de estudantes que acha satisfatório o funcionamento das disciplinas é igual à aproximação que encontrou na alínea (a) dividido por 7038? Justifique a sua resposta.

(d) Em vez do procedimento de amostragem descrito acima decide-se escolher 14 disciplinas, 4 das quais entre as 40 disciplinas pequenas e as 10 restantes entre a disciplinas grandes. Em seguida efectuam-se amostragens aleatórias simples com os tamanhos indicados na alínea (a). A aproximação para o total e respectiva variância sofrem alguma alteração? Qual? Calcule os novos valores.

(e) Tendo em conta os resultados descritos na alínea (a), como aconselharia a planear uma estratificação apenas sobre as 10 disciplinas grandes para obter uma amostra com tamanho final 80 (o que corresponde a cerca de 10% do tamanho das 10 disciplinas grandes registadas).

5. Num $\Pi\text{PS}(N, n, \mathbf{p})$ a média da amostra $\overline{y} = \frac{1}{n}\sum_{i\in s} Y_i$ é um estimador não enviesado para a média da população. Mostre que $\mathbf{p} = \left(\frac{1}{N}, \ldots, \frac{1}{N}\right)$.

6. Ao executar um $\Pi\text{PS}(10, 3, (0.08, 0.12, 0.08, 0.08, 0.12, 0.08, 0.12, 0.12, 0.08, 0.12))$ seleccionou-se a amostra (3,7,9) a que correspondem os seguintes valores para a variável de interesse: (2,-3,5). Calcule uma aproximação para a média da população. Parece-lhe possível construir uma aproximação para a variância da estimativa anterior? Justifique.