Departamento de Matemática - Universidade de Coimbra

2.^{a} Frequência de Amostragem e Sondagens – 2007/08

3/6/08 tempo: 1h30m

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. Uma região geográfica subdivide-se 9 subregiões, $R_1, \ldots, R_9$, conforme se indica na figura ao lado. Nesta figura indicam-se também o número de unidades recenseadas em cada subregião na *frame* que serve de base ao processo de amostragem, correspondendo a uma população com tamanho $N = 882$. Efectuou-se uma amostragem recolhendo duas leituras, designadas por Y e p, por cada unidade incluída na amostra. A variável Y assume valores reais não negativos, enquanto que a variável p é binária, isto é, só assume os valores 0 ou 1. Recolheu-se uma amostra com 100 unidades.

<table>
<tr><td colspan="4">R_1
105</td><td colspan="4">R_2
98</td><td colspan="4">R_3
150</td></tr>
<tr><td colspan="3">R_4
72</td><td colspan="3">R_5
84</td><td colspan="3">R_6
67</td><td colspan="3">R_7
79</td></tr>
<tr><td colspan="6">R_8
123</td><td colspan="6">R_9
104</td></tr>
</table>

Os resultados disponíveis apresentam o valor médio e variância $\overline{y}_{R_i}$, $s^2_{R_i}$ por região relativamente à variável Y, e o valor médio ou proporção $\overline{p}_{R_i}$ por região, relativamente à variável p, conforme a tabela seguinte (indicam-se também o número de unidades de cada subregião que foram incluídas na amostra observada).

i	1	2	3	4	5	6	7	8	9
n_{R_i}	12	11	17	8	9	8	9	14	12
$(\overline{y}_{R_i}, s^2_{R_i})$	(5.75, 4.6)	(1.9, .79)	(2.6, 1.7)	(8.6, 18.1)	(6.1, 5.7)	(.8, .18)	(4.2, 3.9)	(1.7, .7)	(2.2, 1.1)
$\overline{p}_{R_i}$	.75	.74	.22	.89	.29	.91	.11	.86	.87

(a) Não dispondo de mais informações sobre o plano de amostragem, assume-se que a amostra foi seleccionada executando um único SRS$(822, 100)$ sobre a população inteira.

i. Construa os parâmetros média da amostra $\overline{y}$ e variância da amostra $s^2_{\mathbf{Y}}$ relativos à variável Y. (Nota: Caso não responda a esta questão considere, nas alíneas que se seguem, que $\overline{y} = 3.75$ e $s^2_{\mathbf{Y}} = 7.25$.)

ii. Indique um intervalo com confiança 0.95 para a média da população. (Os percentis habituais para a distribuição $\mathcal{N}(0, 1)$ são $z_{0.025} = 1.96$, $z_{0.05} = 1.64$.)

iii. Qual o tamanho de amostra que aconselharia para amostragens futuras para obter um intervalo com confiança 0.95 para a média cuja amplitude fosse inferior ou igual a 0.5?

(b) Suponha agora que a amostra foi obtida efectuando uma estratificação pelas 9 subregiões com os tamanhos indicados na tabela acima pela linha correspondente a n_{R_i}.

i. Calcule uma estimativa não enviesada para a média da população. Explique a diferença relativamente à estimativa encontrada em (a) i.

ii. Construa uma estimativa para a variância do estimador para a média da população.

iii. Obtenha uma estimativa para o valor médio da variável p e respectiva variância.

(c) Com as informações disponíveis, como aconselharia a efectuar a estratificação numa futura amostragem à mesma população, considerando que a variável mais significativa é Y?

V.S.F.F.

(d) Define-se um novo plano de amostragem procedendo da seguinte forma: definem-se os conjuntos de unidades amostrais $\mathcal{C}_1 = R_1 \cup R_2 \cup R_3$, $\mathcal{C}_2 = R_4 \cup R_5 \cup R_6 \cup R_7$ e $\mathcal{C}_3 = R_8 \cup R_9$,

- escolhem-se alguns destes conjuntos de acordo com as seguintes probabilidades

$$\mathrm{P}((\mathcal{C}_1, \mathcal{C}_2)) = 0.5, \qquad \mathrm{P}((\mathcal{C}_2, \mathcal{C}_3)) = 0.3, \qquad \mathrm{P}((\mathcal{C}_2)) = 0.2;$$

- nos conjuntos $\mathcal{C}_1$ e $\mathcal{C}_3$, caso tenham sido seleccionados incluem-se todas as unidades; no conjunto $\mathcal{C}_2$ executa-se um plano de amostragem estratificado pelas quatro subregiões R_4, R_5, R_6 e R_7 recolhendo 8, 9, 8 e 9 unidades respectivamente.

Os totais, relativamente à variável p nos conjuntos $\mathcal{C}_1$ e $\mathcal{C}_3$ são 191 e 195, respectivamente. Utilize a lista de números aleatórios para efectuar a escolha dos conjuntos.

i. Que estimativa constrói para o total da população relativamente à variável p, assumindo que em $\mathcal{C}_2$ se efectuou uma amostragem com os resultados indicados na tabela acima? É não enviesada?

ii. Construa uma estimativa não enviesada para a variância da aproximação obtida na alínea anterior.

2. Considere, na população $\mathcal{U} = \{1, 2, 3, 4, 5\}$ o seguinte plano de amostragem:

S	(1,2)	(1,3)	(1,4)	(1,5)	(2,3)	(2,4)	(2,5)	(3,4)	(3,5)	(4,5)
P	.012	.235	.099	.099	.408	.012	.012	.012	p_1	p_2

Suponha que $\mathbf{Y} = (-3, 2, 12, 0, -1)$.

(a) Este plano de amostragem é um $\Pi\mathrm{PS}(9, 2, \mathbf{p})$ com $\mathbf{p} = (0.2225, 0.222, 0.3335, 0.111, 0.111)$. Identifique p_1 e p_2. (Nota: Caso não responda a esta questão assuma, nas alíneas seguintes, que $p_1 = 0.015$ e $p_2 = 0.006$.)

(b) Utilize a lista de números aleatórias para efectuar a escolha de uma amostra.

(c) Construa estimativas para o total e para a variância da população.

(d) A estimativa para a variância é não enviesada? Justifique.

(e) Construa, caso lhe pareça possível, um intervalo com confiança 95% para a média da população. (Os percentis habituais para a distribuição $\mathcal{N}(0, 1)$ são $z_{0.025} = 1.96$, $z_{0.05} = 1.64$.)