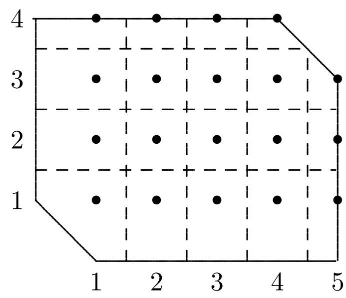
Departamento de Matemática - Universidade de Coimbra
Frequência de Amostragem e Sondagens – 2007/08 – Proposta de resolução

1. (a) Na figura ao lado delimitam-se as zonas que ficam mais próximas de cada ponto de coordenadas naturais. Como se efectua uma escolha ao acaso na região, basta calcular as áreas de cada região e dividir pela área total. A área total da região mede 19. O plano de amostragem fica então definido por



S	$(1,1)$	$(2,1)$	$(3,1)$	$(4,1)$	$(5,1)$	$(1,2)$	$(2,2)$	$(3,2)$	$(4,2)$	$(5,2)$
P	$\frac{1.75}{19}$	$\frac{1.5}{19}$	$\frac{1.5}{19}$	$\frac{1.5}{19}$	$\frac{0.75}{19}$	$\frac{1.5}{19}$	$\frac{1}{19}$	$\frac{1}{19}$	$\frac{1}{19}$	$\frac{0.5}{19}$
S	$(1,3)$	$(2,3)$	$(3,3)$	$(4,3)$	$(5,3)$	$(1,4)$	$(2,4)$	$(3,4)$	$(4,4)$	
P	$\frac{1.5}{19}$	$\frac{1}{19}$	$\frac{1}{19}$	$\frac{1}{19}$	$\frac{3}{8\times 19}$	$\frac{0.75}{19}$	$\frac{0.5}{19}$	$\frac{0.5}{19}$	$\frac{3}{8\times 19}$	

(b) É preciso construir a função de probabilidades acumuladas correspondente à ordenação considerada na alínea anterior (qualquer outra ordenação poderia ser considerada, esta escolha deve-se apenas a razões de ordem prática)

S	$(1,1)$	$(2,1)$	$(3,1)$	$(4,1)$	$(5,1)$	$(1,2)$	$(2,2)$	$(3,2)$	$(4,2)$	$(5,2)$
P	$\frac{1.75}{19}$	$\frac{1.5}{19}$	$\frac{1.5}{19}$	$\frac{1.5}{19}$	$\frac{0.75}{19}$	$\frac{1.5}{19}$	$\frac{1}{19}$	$\frac{1}{19}$	$\frac{1}{19}$	$\frac{0.5}{19}$
	$\frac{1.75}{19}$	$\frac{3.25}{19}$	$\frac{4.75}{19}$	$\frac{6.25}{19}$	$\frac{7}{19}$	$\frac{8.5}{19}$	$\frac{9.5}{19}$	$\frac{10.5}{19}$	$\frac{11.5}{19}$	$\frac{12}{19}$
S	$(1,3)$	$(2,3)$	$(3,3)$	$(4,3)$	$(5,3)$	$(1,4)$	$(2,4)$	$(3,4)$	$(4,4)$	
P	$\frac{1.5}{19}$	$\frac{1}{19}$	$\frac{1}{19}$	$\frac{1}{19}$	$\frac{3}{8\times 19}$	$\frac{0.75}{19}$	$\frac{0.5}{19}$	$\frac{0.5}{19}$	$\frac{3}{8\times 19}$	
	$\frac{13.5}{19}$	$\frac{14.5}{19}$	$\frac{15.5}{19}$	$\frac{16.5}{19}$	$\frac{16.875}{19}$	$\frac{17.625}{19}$	$\frac{18.125}{19}$	$\frac{18.625}{19}$	1	

Escolhe-se agora um número ao acaso em $[0,1]$: $u = 0.972208$. Como $\frac{18.125}{19} = 0.953947 < u \leq \frac{18.625}{19} = 0.980263$, selecciona-se a amostra $s = (3,4)$.

2. (a) $q_1(i) = p_i$, $q_2(i) = 0$ se $i \leq \frac{N}{2}$, $q_2(i) = 1$ se $i > \frac{N}{2}$, $q_2(s) = 0$ para $s \neq (i)$, $q_3(j|i) = \frac{p_j}{1-p_i}$, para $i > \frac{N}{2}$ e $j \neq i$.

(b) O suporte do plano de amostragem é $S = \Big\{(i),\, i \leq \frac{N}{2},\, (i_1,i_2),\, i_1 > \frac{N}{2},\, i_2 \neq i_1\Big\}$. A distribuição de probabilidade é caracterizada por

$$\mathrm{P}(i) = p_i \text{ se } i \leq \frac{N}{2}, \qquad \mathrm{P}(i_1,i_2) = p_{i_1} \times \frac{p_{i_2}}{1-p_{i_1}} \text{ se } i_1 > \frac{N}{2},\, i_2 \neq i_1$$

(c) $\pi_i = p_i + \displaystyle\sum_{j=\frac{N}{2}+1}^{N} \frac{p_i p_j}{1-p_j}$, para todo o $i \in \mathcal{U}$.

$$\pi_{i,j} = \begin{cases} \dfrac{p_i p_j}{1-p_j} & \text{se } i \leq \frac{N}{2},\ j > \frac{N}{2} \\[2ex] \dfrac{p_i p_j}{1-p_i} & \text{se } i > \frac{N}{2},\ j \leq \frac{N}{2} \\[2ex] \dfrac{p_i p_j}{1-p_j} + \dfrac{p_i p_j}{1-p_i} & \text{se } i,\ j > \frac{N}{2} \\[2ex] 0 & i, j \leq \frac{N}{2}. \end{cases}$$

3. (a) A função de probabilidades acumuladas é

1	2	3	4
0.1	0.2	0.3	0.4
0.1	0.3	0.6	1

Para seleccionar três unidades necessitamos de três números ao acaso no intervalo $[0,1]$: $u_1 = 0.028817$, $u_2 = 0.837811$, $u_3 = 0.618123$. Segue-se então que $i_1 = 1$, $i_2 = 4$ e $i_3 = 4$. Isto é, seleccionamos a amostra $s = (1,4,4)$.

(b)

S	(1)	(2)	(3)	(4)		
P	0.1^3	$0.2^3 = 0.008$	$0.3^3 = 0.027$	$0.4^3 = 0.064$		
S	$(1,2)$	$(1,3)$	$(1,4)$	$(2,3)$	$(2,4)$	$(3,4)$
P	$3\times(0.1^2\times 0.2+0.1\times 0.2^2) = 0.018$	0.036	0.060	0.090	0.144	0.252
S	$(1,2,3)$	$(1,2,4)$	$(1,3,4)$	$(2,3,4)$		
P	$6\times 0.1\times 0.2\times 0.3 = 0.036$	0.048	0.074	0.144		

(c) Calculemos detalhadamente a expressão para o estimador simetrizado para a amostra $(1,1,2)$. A classe de equivalência desta amostra é

$$[(1,1,2)] = \{(1,1,2),\ (1,2,1),\ (2,1,1),\ (1,2,2),\ (2,1,2),\ (2,2,1)\}.$$

Nas três primeiras amostras desta classe de equivalência o estimador considerado toma o valor $\frac{10Y_1+5Y_2}{3}$, enquanto que nas três últimas toma o valor $\frac{5Y_1+10Y_2}{3}$. Para calcular o estimador simetrizado ℓ^*, fazemos

$$\begin{aligned}
\ell^*\Big((1,1,2),\mathbf{Y}\Big) &= \frac{\sum_{s'\in[(1,1,2)]}\ell(s',\mathbf{Y})\mathrm{P}(s')}{\sum_{s'\in[(1,1,2)]}\mathrm{P}(s')} \\
&= \frac{\frac{10Y_1+5Y_2}{3}\times 3\times 0.1^2\times 0.2 + \frac{5Y_1+10Y_2}{3}\times 3\times 0.1\times 0.2^2}{3\times(0.1^2\times 0.2+0.1\times 0.2^2)} \\
&= \frac{2Y_1+2.5Y_2}{0.9}.
\end{aligned}$$

Este cálculo seria repetido, obtendo-se o mesmo resultado para qualquer outra amostra pertencente a $[(1,1,2)]$. Assim, para caracterizar completamente o estimador simetrizado ℓ^* basta indicar o valor que este assume para cada amostra no plano reduzido e simetrizado. O cálculo explícito das expressões que se indicam abaixo obtém-se repetindo os argumentos anteriores para cada classe de equivalência.

S	(1)	(2)	(3)	(4)		
ℓ^*	$5Y_1$	$5Y_2$	$5Y_3$	$5Y_4$		
S	$(1,2)$	$(1,3)$	$(1,4)$	$(2,3)$	$(2,4)$	$(3,4)$
ℓ^*	$\frac{2Y_1+2.5Y_2}{0.9}$	$\frac{2.5Y_1+3.5Y_3}{1.2}$	$\frac{3Y_1+4.5Y_4}{1.5}$	$\frac{3.5Y_2+4Y_3}{1.5}$	$\frac{4Y_2+5Y_4}{1.8}$	$\frac{5Y_3+5.5Y_4}{2.1}$
S	$(1,2,3)$	$(1,2,4)$	$(1,3,4)$	$(2,3,4)$		
ℓ^*	$\frac{5}{3}(Y_1+Y_2+Y_3)$	$\frac{5}{3}(Y_1+Y_2+Y_4)$	$\frac{5}{3}(Y_1+Y_3+Y_4)$	$\frac{5}{3}(Y_2+Y_3+Y_4)$		

4. Na resolução deste problema consideramos apenas a versão reduzida do plano de amostragem.

(a) Existem $\binom{9}{3} = 84$ amostras possíveis. Como há 4 a que atribuímos probabilidade p, haverá 80 a que atribuímos probabilidade $2p$. Assim, $4p + 80 \times 20 = 164p = 1$, ou seja $p = \frac{1}{164}$.

(b) Seja $u = 0.205081$ um número escolhido ao acaso no intervalo $[0, 1]$. Consideramos as 84 amostras possíveis ordenadas por ordem lexicográfica crescente: $(1, 2, 3) < (1, 2, 4) < \cdots < (1, 2, 9) < (1, 3, 4) < \cdots < (1, 8, 9) < (2, 3, 4) < \cdots$. Com a unidade 1 na primeira posição há $\binom{8}{2} = 28$ amostras. Destas uma tem probabilidade $\frac{1}{164}$, enquanto que as outras 27 têm probabilidade $\frac{2}{164}$. Assim, a probabilidade acumulada até à amostra $(1, 8, 9)$ é $\frac{1}{164} + \frac{27\times2}{164} = \frac{55}{164} = 0.335366$. Isto é, a amostra a seleccionar tem como primeira unidade a região 1. Entre estas amostras, as probabilidade acumuladas são sucessivamente $\frac{2}{163}$, $\frac{3}{164}$, $\frac{5}{164}$, $\frac{7}{164}$,$\ldots$, $\frac{55}{164}$. Como $\frac{33}{164} = 0.20122 < u \leq \frac{35}{164} = 0.213415$, procuramos a amostra a que corresponde a probabilidade acumulada $\frac{35}{164}$, ou seja, a amostra $s = (1, 4, 9)$.

(c) $\pi_1 = \pi_3 = \pi_7 = \pi_9 = \frac{1}{164} + \left(\binom{8}{2} - 1\right) \times \frac{2}{164} = \frac{55}{164}$, $\pi_2 = \pi_4 = \pi_6 = \pi_8 = 2 \times \frac{1}{164} + 26 \times \frac{2}{164} = \frac{27}{82}$, $\pi_5 = 28 \times \frac{2}{164} = \frac{14}{41}$.

(d) Como $\sum_{s:1\in s} \ell_{s,1}\mathrm{P}(s) = \frac{1}{3} \times \pi_1 = \frac{55}{492} = 0.111789 \neq 1$, o estimador é enviesado para o total da população.

(e) Há que escolher as constantes por forma a que se cumpra a condição necessária e suficiente de não enviesamento para o total: $\sum_{s:i\in s} \ell^*_{s,i}\mathrm{P}(s) = 1$, $i = 1, \ldots, 9$, onde $\ell^*_{s,i}$ designa os coeficientes que caracterizam a representação linear do estimador ℓ^*. Os coeficientes de $\ell^*_{s,i}$ são sempre iguais a $\frac{1}{3}$ para todas as amostras s distintas de $(1, 8, 9)$, $(2, 8, 9)$,$\ldots$, $(7, 8, 9)$. Nas primeiras seis destas amostras apenas se altera o coeficiente associado à primeira unidade. Há, portanto, que ter atenção às probabilidades de selecção de cada amostra. No caso de $i = 1$ somos então conduzidos à seguinte equação:

$$\sum_{s:i\in s} \ell^*_{s,i}\mathrm{P}(s) = \frac{1}{3} \times \frac{1}{164} + 26 \times \frac{1}{3} \times \frac{2}{164} + a_1 \times \frac{2}{164} = 1 \Longleftrightarrow a_1 = \frac{129}{2}.$$

No caso $i = 3$ encontramos a mesma equação, pelo que $a_3 = \frac{129}{2}$. No casos $i = 2$, 4, encontramos

$$\sum_{s:i\in s} \ell^*_{s,i}\mathrm{P}(s) = 2 \times \frac{1}{3} \times \frac{1}{164} + 25 \times \frac{1}{3} \times \frac{2}{164} + a_i \times \frac{2}{164} = 1 \Longleftrightarrow a_2 = a_4 = \frac{220}{3}.$$

Para $i = 5$,

$$\sum_{s:i\in s} \ell^*_{s,i}\mathrm{P}(s) = 27 \times \frac{1}{3} \times \frac{2}{164} + a_5 \times \frac{5}{164} = 1 \Longleftrightarrow a_5 = 73.$$

Para $i = 6$,

$$\sum_{s:i\in s} \ell^*_{s,i}\mathrm{P}(s) = 27 \times \frac{1}{3} \times \frac{2}{164} + a_6 \times \frac{1}{164} = 1 \Longleftrightarrow a_6 = 146.$$

Para $i = 7$,

$$\sum_{s:i\in s} \ell^*_{s,i}\mathrm{P}(s) = \frac{1}{3} \times \frac{1}{164} + 26 \times \frac{1}{3} \times \frac{2}{164} + a_7 \times \frac{2}{164} = 1 \Longleftrightarrow a_7 = \frac{129}{2}.$$

Finalmente, para $i = 8,\ 9$,

$$\sum_{s:i\in s} \ell^*_{s,i}\mathrm{P}(s) = \frac{1}{3} \times \frac{1}{164} + 26 \times \frac{1}{3} \times \frac{2}{164} + a_i \times \frac{2}{164} = 1 \Longleftrightarrow a_8 = a_9 = \frac{439}{6}.$$