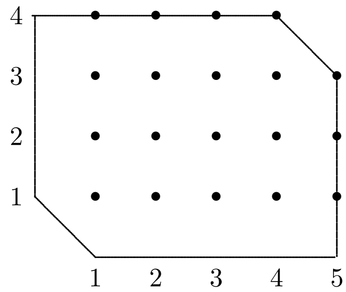
Departamento de Matemática - Universidade de Coimbra
Frequência de Amostragem e Sondagens – 2007/08

3/4/08 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. Considere a região indicada ao lado, onde estão assinalados com o símbolo $\bullet$ os pontos de coordenadas naturais.



   (a) Descreva o plano de amostragem na população $\mathcal{U} = \{1, 2, 3, 4, 5\}$ obtido seleccionando a amostra correspondente ao ponto de coordenadas naturais que fica mais próximo do ponto que for seleccionado na região dada.

   (b) Utilize a lista de números aleatórios para escolher uma amostra de acordo com plano de amostragem descrito.

2. Considere-se uma população $\mathcal{U}$ com N unidades, em que N é par, e uma distribuição de probabilidade sobre $\mathcal{U}$ descrita por $p_1, \dots, p_N$. Procedemos à escolha de unidades em $\mathcal{U}$ da seguinte forma:

   i. escolhe-se a primeira unidade, i_1, a incluir na amostra de acordo com a distribuição de probabilidade dada;

   ii. se $i_1 \leq \frac{N}{2}$ pára-se a construção da amostra e ficamos com $s = (i_1)$;

   iii. caso $i_1 > \frac{N}{2}$ escolhemos uma segunda unidade, i_2 distinta de i_1, para a amostra de acordo com a distribuição de probabilidade $p_j^* = \frac{p_j}{1-p_{i_1}}$, $j \in \mathcal{U}$, $j \neq i_1$, e tomamos $s = (i_1, i_2)$.

   (a) Descreva o algoritmo de escolha aleatória definido.

   (b) Descreva o plano de amostragem.

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

3. Considere-se na população $\mathcal{U} = \{1, 2, 3, 4\}$ a distribuição de probabilidade (0.1, 0.2, 0.3, 0.4). Procede-se à selecção de três unidades da população escolhendo três vezes uma unidade em $\mathcal{U}$, recolocando sempre a unidade seleccionada, de acordo com a distribuição dada. Esta escolhe permite, portanto, que haja repetições de unidades na amostra.

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

   (b) Descreva a versão reduzida e simetrizada do plano de amostragem.

   (c) Relativamente ao plano descrito (que admite repetições), consideramos o estimador $\ell\Big((i, j, k), \mathbf{Y}\Big) = \frac{5}{3}(Y_i + Y_j + Y_k)$. Descreva o estimador simetrizado correspondente.

4. Pretende-se efectuar uma amostragem sobre um conjunto de 9 regiões dispostas como na figura ao lado. Decidiu-se construir amostras de dimensão 3. A região 5 é considerada central e todas as outras consideradas periféricas. Acredita-se que estas regiões periféricas apresentam comportamentos semelhantes, no que à variável de interesse diz respeito, quando são vizinhas próximas. Assim, pretende-se construir um plano de amostragem que permita a escolha de qualquer amostra com três unidades, mas dê menos peso à possibilidade de escolha das amostras contendo as unidades de um dos seguintes conjuntos: $\{1, 2, 4\}$, $\{2, 3, 6\}$, $\{4, 7, 8\}$ ou $\{6, 8, 9\}$. A estas amostras atribui-se a probabilidade p, enquanto que a todas as outras amostras possíveis se atribui a probabilidade $2p$.

1	2	3
4	5	6
7	8	9

V.S.F.F.

(a) Calcule p, especificando se se refere ao plano de amostragem reduzido ou não.

(b) Utilize a lista de números aleatórios para seleccionar uma amostra de acordo com plano de amostragem descrito. (Caso não tenha resolvido a alínea anterior assuma que $p = 0.006$.)

(c) Calcule as probabilidades de inclusão de primeira ordem.

(d) Para efeitos de inferência sobre o total da população alguém havia programado cálculo com o estimador $\ell\Big((i_1, i_2, i_3), \mathbf{Y}\Big) = \frac{1}{3}\left(Y_{i_1} + Y_{i_2} + Y_{i_3}\right)$. Justifique que o estimador é enviesado para o total da população.

(e) Para corrigir o enviesamento de ℓ propomo-nos construir um estimador ℓ^*, alterando um número mínimo de coeficientes, caracterizado da seguinte forma (assume-se que tratamos a versão reduzida do plano de amostragem):

$$\ell^*\Big((i_1, i_2, i_3), \mathbf{Y}\Big) = \begin{cases} a_{i_1} Y_{i_1} + \frac{1}{3} Y_{i_2} + \frac{1}{3} Y_{i_3} & \text{se } i_1 = 1, \ldots, 6,\ i_2 = 8,\ i_3 = 9 \\ a_7 Y_7 + a_8 Y_8 + a_9 Y_9 & \text{se } s = (7, 8, 9) \\ \ell(s, \mathbf{Y}) & \text{caso } s \notin \{(1,8,9),\ (2,8,9), \ldots, (7,8,9)\}. \end{cases}$$

Calcule as constantes $a_1, \ldots, a_7, a_8, a_9$ por forma que ℓ^* seja, de facto, não enviesado para o total da população.