Universidade Federal de Lavras
Departamento de Ciências Exatas
Prof. Daniel Furtado Ferreira

$5^{\underline{a}}$ Aula Prática **Amostragem**

1) Uma população é formada de $N = 35$ árvores de uma determinada espécie, pertencentes a um parque ecológico, que possuem os seguintes diâmetros a altura do peito em cm (DAP): 25, 20, 35, 21, 22, 22, 24, 25, 30, 38, 24, 20, 21, 25, 20, 15, 25, 23, 20, 24, 28, 24, 24, 22, 28, 26, 23, 19, 22, 27, 25, 23, 28, 27, 42. Com o objetivo de estimar o DAP (diâmetro a altura do peito) médio, como você extrairia uma amostra simples ao acaso, de tamanho $n = 10$ desta população? Dê todos os detalhes e estime a média. Compare com a média da população, determinando o erro relativo de estimação percentual por: $er = (\bar{X} - \mu)/\mu \times 100\%$. Quantas amostras de tamanho $n = 10$ podemos extrair desta pequena população, considerando amostragem com reposição e sem reposição? Dê sua opinião sobre estes valores.

2) Qual é a principal diferença entre amostra probabilística e não probabilística?

3) Os dados apresentados a seguir referem-se às variações de pesos corporais em $N = 20$ ratos em g/animal. Os dados foram avaliados em raças endogâmicas pequenas de ratos e em fêmeas, com o objetivo de fazer uma caracterização genética. Supondo que as $N = 20$ fêmeas constituam toda a população, para fins de treinamento, faça amostras de 10%, 20%, 30%, 50% e 60% do tamanho populacional e estime o erro (%) em cada caso por $er = (\bar{X} - \mu)/\mu \times 100\%$ para o peso das fêmeas. Comente sobre os resultados obtidos. Plote n na abscissa versus o erro relativo na ordenada e discuta os resultados obtidos.

Peso de ratos em g (fêmeas)			
15,77	21,47	19,17	17,40
17,76	21,65	17,90	20,38
14,45	22,71	17,51	15,37
19,19	17,93	19,62	19,87
18,47	19,14	15,99	22,30

4) Uma empresa agrícola tem $N = 3.414$ empregados subdivididos nos seguintes setores:

Setores (h)	Número de funcionários (N_h)
Administrativo	314
Transporte	948
Campo	1.451
Outros	701
Total	3.414

Para se estudar o nível salarial médio da empresa, resolveu-se fazer uma amostra de $n = 180$ funcionários. Você julga que a ASA, seria apropriada, para este caso? Se não for, o que você recomendaria? Dê todos os detalhes do dimensionamento da amostra.

5) Se na amostra, do exercício anterior, as médias em $L = 4$ estratos forem dadas por:

Setores (h)	Número de funcionários amostrais(n_h)	($\bar{X}_h$)
Administrativo	$n_1 =$	2.545,00
Transporte	$n_2 =$	480,00
Campo	$n_3 =$	680,00
Outros	$n_4 =$	987,00
Total	$n = 180$	

Estimar a média da população utilizando os dois métodos apresentados em aula, considerando o dimensionamento de amostra obtido no exercício anterior.

6) Diferencie amostra simples ao acaso e amostra sistemática.

7) Em que situação você recomendaria utilizar a ASA ou a amostra sistemática em substituição a um processo de amostragem estratificada. Justifique sua resposta

8) Qual é a principal idéia por trás da determinação do tamanho amostral de um determinado estrato na amostragem estratificada ótima?

9) Diferenciar amostragem estratificada uniforme e amostragem estratificada proporcional, indicando a principal característica no dimensionamento da amostra para um determinado estrato neste último processo.

Gabarito

1) Podemos extrair uma amostra de tamanho $n = 10$, sem reposição, da seguinte forma: a) enumerando a população de 1 a $N = 35$ e sorteando $n = 10$ números aleatórios entre 1 e 35. Se algum destes números se repetir, sorteamos outro número. Estes números representam as $n = 10$ árvores sorteadas. Registramos seu DAP para formamos nossa amostra. Convém enfatizar que em uma situação real, temos apenas o diâmetro das árvores que foram amostradas. Neste exemplo temos todos os DAPs, pois é um exemplo didático. Sorteamos um número # da árvore, da seguinte forma: $\#\text{Árvore} = trunc(Random \times N) + 1$, em que $trunc$ retorna a parte inteira do argumento;

   Fizemos isso para o exemplo, considerando $n = 10$ e obtivemos a seguinte amostra: 15, 22, 21, 25, 19, 25, 20, 21, 24, 26. Cada aluno, utilizando um processo aleatório de sorteio irá produzir uma amostra diferente. A média desta amostra é dada por: $\bar{X} = (15 + \ldots + 26)/10 = 21{,}8$. A média da população é $\boldsymbol{\mu} = (25 + 20 + \ldots + 42)/35 = 24{,}77143$.

   Assim, o erro relativo foi:

$$
\begin{aligned}
er =& \frac{\bar{X} - \mu}{\mu} \times 100 = \frac{21{,}8 - 24{,}77143}{24{,}77143} \times 100 \\
=& -11{,}99539\%.
\end{aligned}
$$

   Assim, erramos para menos 11,9954%, ou seja, nossa amostra subestimou a média da população.

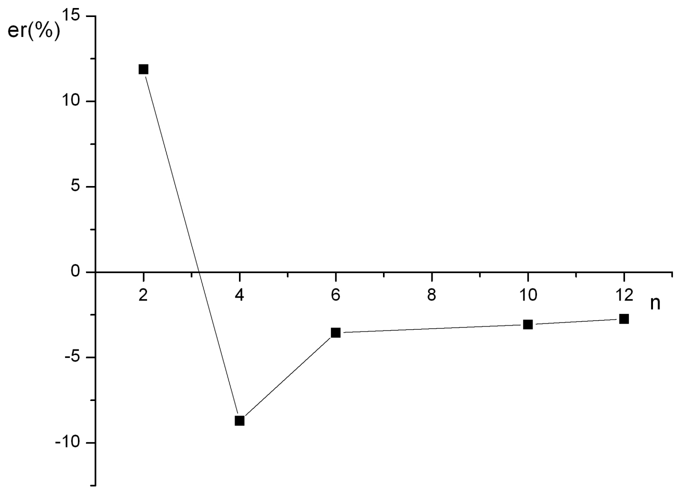
   O número possível de amostras de tamanho $n = 10$, sem reposição, dessa população é dado por $\binom{N}{n} = \binom{35}{10} = 183.579.396$. Podemos observar que o número de amostras de tamanho $n = 10$ extraída sem reposição de uma população de tamanho $N = 35$ é muito grande, ou seja, de aproximadamente 184 milhões de possibilidades. Com reposição esse número é de $N^n = 35^10 = 2{,}76 \times 10^{15}$. Assim, em populações reais, este número é extremamente grande, refletindo a complexidade dos processos de amostragem.

2) Amostragem probabilística é aquela em que todos elementos da população possuem probabilidade não-nula de participar da amostra e sua principal característica é o uso do sorteio. Se por algum razão algum elemento ou grupo de elementos da população possuir probabilidade nula de participar da amostra ou a amostragem for feita sem sorteio, então a amostragem é considerada não-probabilística.

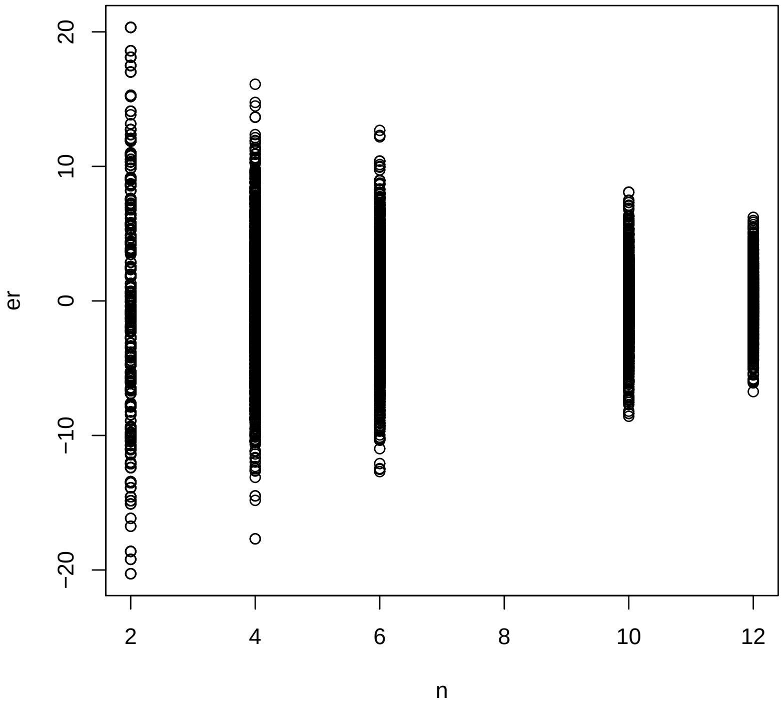
3) A média populacional do peso de ratos é: $\mu = (15{,}77 + 21{,}47 + \ldots + 22{,}30)/20 = 18{,}7025$. Para obtermos os erros relativos a cada tamanho de amostra, construímos a seguinte tabela contendo as amostras sorteadas de cada tamanho (sem reposição) e a estimativa da média. Cada amostragem foi feita conforme procedimento descrito no exercício resolvido 1.

Tamanho (n)	Amostra	$\bar{X}$	$er\%$
2	21,47; 20,38	20,9250	11,88%
4	15,99; 19,14; 15,77; 17,40	17,0750	−8,70%
6	17,93; 19,87; 17,40; 19,19; 15,37; 18,47	18,0383	−3,55%
10	17,93; 17,51; 17,40; 18,47; 17,90; 14,45; 15,37; 22,71; 20,38; 19,17	18,1290	−3,07%
12	15,77; 19,62; 22,30; 14,45; 22,71; 18,47; 19,14; 15,99; 19,17; 15,37; 17,90; 17,40	18,1908	−2,74%

   O gráfico correspondente é dado por:



Verificamos que existe uma “tendência” do valor absoluto do erro decrescer na medida que o tamanho da amostra n aumenta. Fizemos 1.000 repetições deste procedimento em um programa de análise estatística e o resultado gráfico é dado por:



Observamos que existe uma tendência das amplitudes dos erros relativos reduzirem com o aumento do tamanho da amostra n.

4) Como a população possui, possivelmente, uma heterogeneidade de salários entre os diferentes estratos a ASA não é apropriada. Devemos fazer uma amostragem estratificada, que, no caso, deve ser a AE proporcional. A amostra deve ser dimensionada em cada estrato considerando seu tamanho, ou seja, quanto maior o estrato populacional, maior deve ser a amostra naquele estrato. O dimensionamento segue a seguinte expressão: $n_h = n \times N_h/N$. Assim, para o primeiro estrato temos: $n_1 = n \times N_1/N = 180 \times 314/3.414 = 16{,}55 \approx 17$. Para os demais estratos, utilizamos esta fórmula e obtivemos os seguintes resultados

Setores (h)	Número de funcionários (N_h)	n_h
Administrativo	314	17
Transporte	948	50
Campo	1.451	77
Outros	701	37
Total	$N = 3.414$	$n = 181$

Como todos os arredondamentos foram feitos para cima, então a amostra efetiva deverá ser de $n = 181$. Para amostrarmos cada estrato, podemos utilizar tanto uma ASA como uma amostragem sistemática (AS). Para aplicar uma ASA é necessário ter estratos enumeráveis e para uma AS, devemos possuir algum tipo de distribuição espacial do estrato de forma a permitir um processo de saltos regulares entre os seus elementos.

5) Podemos utilizar como peso os tamanhos amostrais ou os tamanho populacionais de cada estrato. A tabela completa com os pesos e tamanho de amostra retificado pelos ajustes feitos no exercício anterior são:

Setores (h)	N_h	n_h	$(\bar{X}_h)$
Administrativo	314	$n_1 = 17$	2.545,00
Transporte	948	$n_2 = 50$	480,00
Campo	1.451	$n_3 = 77$	680,00
Outros	701	$n_4 = 37$	987,00
Total	$N = 3.414$	$n = 181$	

Utilizando o primeiro estimador:

$$\bar{X}_{est} = \frac{314 \times 2.545{,}00 + \ldots + 701 \times 987}{3.414} = 859{,}03$$

e aplicando o segundo estimador, temos

$$\bar{X} = \frac{17 \times 2.545{,}00 + \ldots + 37 \times 987}{181} = 862{,}67.$$

Neste caso, os dois estimadores são equivalentes, pois $n_h/n = N_h/N$. A pequena diferença ocorrida se deveu ao arredondamento dos tamanhos dos estratos amostrais.

6) ASA: Amostragem utilizada em situações em que a população é homogênea e enumerável. AS: também utilizada em populações homogêneas, mas que tenha uma distribuição espacial que permite a utilização de um processo de amostragem em que são tomados os elementos sistematicamente (regularmente) de acordo com uma razão $r = N/n$. Assim, os dois tipos de amostragem diferem basicamente na forma em que a amostragem é realizada.

7) Em populações homogêneas, pois se a população for heterogênea e for possível realizar um subdivisão em estratos homogêneos, devemos recomendar a amostragem estratificada.

8) O tamanho do estrato amostral é diretamente proporcional ao tamanho do estrato populacional e a sua variabilidade. A idéia é que quanto maior for estrato amostral, maior deve ser sua amostra para representá-lo adequadamente; da mesma forma e não menos importante, quanto mais variável for o estrato, maior heterogeneidade, maior deve ser a representatividade do estrato. Esta análise é facilmente realizada observando a expressão para dimensionarmos o tamanho do estrato amostral, que é dada por:

$$n_h = \frac{N_h \sigma_h}{\sum_{i=1}^{L} N_h \sigma_h} \times n.$$

9) A amostragem estratificada uniforme deve ser utilizada quando os estratos populacionais possuírem tamanhos aproximadamente iguais e a amostragem estratificada proporcional, quando estes estratos tiverem tamanhos muito diferentes uns dos outros. Na amostragem estratificada proporcional o tamanho do estrato amostral é diretamente proporcional ao tamanho do estrato populacional. Assim, quanto maior N_h maior será o tamanho da amostra naquele estrato (n_h). Da mesma forma, a expressão utilizada para dimensionar o estrato amostral permite que se faça esta análise. A expressão correspondente é:

$$n_h = \frac{N_h}{N} \times n.$$