

Instituto Superior de Agronomia (2019/2020)

Tópicos de uma resolução de exercícios das Aulas Práticas de Amostragem e Análise Ambiental Módulo I

1. (resolvido na aula)
2. (resolvido na aula)
3. Trata-se de uma amostragem aleatória simples sem reposição numa população de dimensão $N = 800$.

$$\text{Tem-se } n = 20 \quad f = 20/800 = 0.025$$

a) Para a espécie preponderante, $\sum_{i=1}^{20} x_i = 916$; $\sum_{i=1}^{20} x_i^2 = 46770$. Estimativas:

i) $\bar{x} = 45.8$ árvores/parcela.

ii) $s_x'^2 = 253.5368$ (árvores/parcela)²

iii) $x_T^* \equiv t^* = N\bar{x} = 800 \times 45.8 = 36640$ árvores.

b) Para estimar o número total de árvores da floresta, é preciso calcular o nº de árvores observado em cada parcela, seja $z_i = x_i + y_i$, $i = 1, \dots, 20$

$$\sum_{i=1}^{20} z_i = 1178 \quad \sum_{i=1}^{20} z_i^2 = 74910$$

$$\bar{z} = 58.9 \text{ árvores/parcela} \quad \text{e} \quad s_z'^2 = 290.8316 \text{ (árvores/parcela)}^2.$$

Então uma estimativa do número total de árvores na floresta é dada por

$$z_T^* = 800 \times 58.9 = 47120 \text{ árvores na floresta.}$$

A precisão do estimador Z_T^* , é dada pela variância ou pelo desvio padrão de Z_T^* .

$Var(Z_T^*) = N^2(1-f)\frac{\sigma_z'^2}{n}$. Como desconhecemos $\sigma_z'^2$, teremos que usar a sua estimativa,

$$s^2(Z_T^*) \equiv \widehat{Var}[Z_T^*] = N^2(1-f)\frac{s_z'^2}{n} = 9073945.$$

$$\text{O erro padrão é} \quad s'(Z_T^*) = N s_z' \sqrt{\frac{1-f}{n}} = 3012.299 \text{ árvores.}$$

c) Um intervalo de confiança a 99% para o número total de árvores na floresta é dado por:

$$\left[N\bar{z} - t_{0.005(19)} N s_z' \sqrt{\frac{1-f}{n}}; N\bar{z} + t_{0.005(19)} N s_z' \sqrt{\frac{1-f}{n}} \right] [=] 38502.01; 55737.99$$

d) $\alpha = 0.05$ e $\epsilon = 3$

Um valor inicial para a dimensão de amostra a recolher é $n_0 \geq \left(\frac{t_{\alpha/2} s'_x}{\epsilon} \right)^2 \approx$
 $\frac{4 \times 253.5368}{9} = 112.6830$

Vamos agora introduzir a correcção de população finita, i.e.,

$n \geq \frac{n_0}{1 + n_0/N} = 98.7708$. Deve então recolher-se uma amostra de dimensão $n = 99$.

e) (*) A razão de que falamos é $R = X_T/Y_T = \mu_X/\mu_Y$.

i) Uma estimativa de R é dada por $r^* = \bar{x}/\bar{y} = x_T^*/y_T^* = 3.496$.

ii) Uma estimativa da variância do estimador R^* é

$$s'^2(R^*) = \frac{(1-f) \sum_{i=1}^n x_i^2 - 2r^* \sum_{i=1}^n x_i y_i + r^{*2} \sum_{i=1}^n y_i^2}{n\bar{y}^2} = 0.1886433$$

4. Ver os comandos do R, que estão no Anexo 1.

5. Dividir o campo rectangular de 5000 m \times 4000 m em parcelas de 100 m \times 100 m, quer dizer que teremos 50 em comprimento e 40 em largura, por exemplo.

O nº total de parcelas é então $N = 2000$. Vamos seleccionar ao acaso $n = 20$

Um programa em R para fazer aquela escolha ao acaso é, por exemplo:

```
#Sejam
#i=1....40,
#j=1...50
n<-20
linhas<-seq(1:40);colunas<-seq(1:50)
i<-sample(linhas,n);i
j<-sample(colunas,n);j
amostra<-cbind(i,j);amostra
```

Procure correr e ver se compreende.

6.

7. **Solo 1** (x_1) $\sum_{j=1}^{16} x_{1j} = 135.7$; $\sum_{j=1}^{16} x_{1j}^2 = 1183.75$

Solo 2 (x_2) $\sum_{j=1}^{40} x_{2j} = 212.1$; $\sum_{j=1}^{40} x_{2j}^2 = 1265.81$

A amostragem foi feita sem reposição, mas pode admitir-se $f \simeq 0$.

a) Estimativa do valor médio de pH e respectiva precisão:

i) Para cada tipo de solo separadamente.

Solo 1: $\bar{x}_1 = 8.48125$; variância $s_{X_1}^{\prime 2} = s_{x_1}^{\prime 2}/n = 0.1368516$ ou o erro padrão de \bar{X}_1 que é $s_{X_1}'/\sqrt{n} = 0.36993$.

Solo 2: $\bar{x}_2 = 5.3025$; variância $s_{X_2}^{\prime 2} = s_{x_2}^{\prime 2}/n = 0.0904806$ ou o erro padrão de \bar{X}_2 que é $s_{X_2}'/\sqrt{n} = 0.3008$.

ii) Para $n = 56$, necessitamos de ter

$$\sum_{j=1}^{56} x_j = \sum_{j=1}^{16} x_{1j} + \sum_{j=1}^{40} x_{2j} = 347.8$$

$$\sum_{j=1}^{56} x_j^2 = \sum_{j=1}^{16} x_{1j}^2 + \sum_{j=1}^{40} x_{2j}^2 = 2449.56$$

$$\bar{x} = 6.2107 \quad \text{erro padrão de } \bar{X} \text{ é } s_x'/\sqrt{n} = 0.30657.$$

iii) Amostra aleatória estratificada, recolhida com afectação proporcional.

Se a afectação é proporcional significa que $f_i = \frac{n_i}{N_i} = \frac{n}{N} \iff \frac{n_i}{n} = \frac{N_i}{N} = W_i$

$$\text{tem-se } W_1 = \frac{16}{56} = 0.2857143 \quad W_2 = \frac{40}{56} = 0.7142857$$

e a estimativa do pH médio é $\bar{x}_{st} = 6.2107$.

A estimativa de $Var[\bar{X}_{st}]$ é $s_{X_{st}}^{\prime 2} = \sum W_i^2(1 - f_i)s_i^{\prime 2}/n_i$, que, considerando de novo $f_i \simeq 0$ dá $s_{X_{st}}^{\prime 2} = 0.05733513$ ou o erro padrão de \bar{X}_{st} que é $s_{X_{st}}' = 0.02394476$.

b) Para construir um I.C. a 95% para o valor médio de pH no caso a) iii), vamos usar aproximadamente a Normal, sendo dado por

$$\left] \bar{x}_{st} - z_{0.025} s_{X_{st}}'; \bar{x}_{st} + z_{0.025} s_{X_{st}}' \right[=]5.7414; 6.6800[$$

c) Sendo fixo o custo de amostragem de cada observação individual e no caso de se pretender recolher $n = 60$, temos para cada estrato $i = 1, 2$

$$n_i = n \frac{W_i \sigma_i'}{\sum_i W_i \sigma_i'}, \text{ com } \sigma_i' \text{ estimado por } s_i' \text{ e usando a (única) informação de}$$

que dispomos quando os σ_i' foram obtidos com a afectação proporcional, i.e., $W_1 = 0.2857143$ e $W_2 = 0.7142857$, de que resulta

$$n_1 = 14 \text{ e } n_2 = 46$$

8. Para este exercício consultar o script pois cada um de vós irá obter uma amostra

a) Construímos um vector com a população toda, com dimensão $N = 588$. Por exemplo

```

pop<-c(small,medio,large);pop

#Amostra que vou usar

>x<-sample(pop,25,rep=F);x
 [1] 37 17 44 33 34 12 23 42 36 48 22 52 41 39 28 12
[17] 31 29 28 29 23 23 48 74 14

```

- b) Estimativa do investimento médio $\bar{x} = 32.76$. Note que o verdadeiro investimento médio/exploração (que neste caso excepcionalmente se consegue saber) é $\mu = 37.09184$ e tem-se ainda $\sigma'^2 = 356.9251$

Estimativa do investimento total $x_T^* = N \times \bar{x} = 19262.88$

I.C para μ e para X_T — Como conhecemos a população podemos usar o verdadeiro valor da variância. Então para μ , por exemplo, calcular: $\bar{x} \pm z_{0.025} \sigma' \sqrt{\frac{1-f}{n}}$

I.C. para μ a 95% de confiança [25.51344; 40.00656]

I.C. para X_T a 95% de confiança [15001.90; 23523.86]

- c) Dimensão da amostra a considerar por forma a estimar o investimento médio com uma tolerância de 0.2 com um risco inferior a 5% ?

$\alpha = 0.05$ e $d = 0.2$

$n_0 \geq \left(\frac{z_{0.025} \sigma'}{d} \right)^2 = 34279.09$!!!! Bem estranho convenhamos, mas com a tolerância que se dá!!!

Mesmo assim, introduzindo a correcção de população finita

$n \geq \frac{n_0}{1 + n_0/N} = 578$... como esperávamos era necessário quase recolher toda a população

Nota: o valor da tolerância é disparatada. Experimentem valores mais razoáveis.

- d) Agora é necessário tirar uma amostra de dimensão $n_i = ?$ em cada estrato. Se pretendemos fazer uma afectação proporcional então

$n_i/N_i = n/N \implies n_i = nN_i/N = nW_i$.

Donde $n_1 = 9$ $n_2 = 11$ e $n_3 = 5$.

- e) O investimento médio por exploração agrícola calcula-se agora como

$\bar{x}_{st} = \sum_i W_i \bar{x}_i$

- f) Sejam $s_1'^2$, $s_2'^2$ e $s_3'^2$ as estimativas da variância em cada estrato.

Temos $d = 0.2$ e $\alpha = 0.05$. esta tolerância refere-se à semi-amplitude do I.C.

para μ , ou seja $d = t_{\alpha/2} \sqrt{s'^2(\bar{x}_{st})} \implies s'^2(\bar{x}_{st}) = \frac{d^2}{t_{\alpha/2}^2} = d^2/4 \implies s'^2(\bar{x}_{st}) = 0.01$, é o que nos slides 85 e 86 está designado por V

Considerando os custos $c_i = c$, fixos, a dimensão da amostra a recolher em cada estrato será dada por

$$n_i \simeq \frac{\sum_i W_i s'_i}{V + \frac{1}{N} \sum_i W_i s'^2_i} W_i s'_i$$

9. $N = 12400$, $n = 20$, $f = 0.0016$, x - n.º de pessoas por apartamento

a) $\bar{x} = 3.3$ pessoas/apartamento.

$$x_T^* \equiv t^* = N\bar{x} = 40920 \text{ pessoas}$$

I.C. a 90% para X_T é]35513.51 ; 46326.49[

b) Seja v_i a variável tomando valores 1 ou 0 consoante no apartamento i vivem mais de 3 pessoas ou não, respectivamente.

$$\text{i) } \sum_i v_i = 8, \quad p^* = 8/20 = 0.4.$$

A precisão do estimativa é dada por $\widehat{Var}[p^*] = (1 - f) \frac{p^* q^*}{n - 1} = 0.012611$

ii) uma estimativa do total de apartamentos com mais de 3 pessoas é
 $v_T^* = 12400 \times 0.4 = 4960$.

c) Seja y o número de automóveis existentes em cada apartamento da amostra.

$$\sum_i y_i = 24.$$

i) Pede-se uma estimativa de $R = X_T/Y_T$, que é $r^* = \frac{\sum x_i}{\sum y_i} = 2.75$

ii) Tem-se $Y_T = 11000$. Usando esta informação pede-se que se calcule agora
 $\bar{x}_R = r^* \mu_y = 2.44$

10. a) Considerando só as maiores herdades $N = 266$, $n = 13$, $\sum_i y_i = 11$, $\sum_i x_i = 873$

y - indica se há ou não trigo x - área de trigo na herdade.

i) $\bar{x} = 67.15$ ha de trigo/herdade

$$x_{tot}^* = 266 \times 67.15 = 17862.92 \text{ ha de trigo nas maiores herdades}$$

I.C. a 95% $17862.92 \pm t_{0.025,12} N_i s'_x \sqrt{\frac{1-f_i}{n_i}}$

ii) $\hat{p} = 11/13 = 0.8462$

iii) Consideremos $\alpha = 0.05$ e $d = 0.01$

Um valor inicial para a dimensão de amostra a recolher é $n_0 \geq \frac{z_{\alpha/2}^2}{4d^2} = 9604$

Vamos agora introduzir a correcção de população finita, i.e.,

$n \geq \frac{n_0}{1 + (n_0 - 1)/N} = 258.8574$. Deve então recolher-se uma amostra de dimensão $n = 259$, o que é um número muito elevado, mas a precisão foi exagerada. Experimente pedir $d = 0.05$, o que ainda é um valor muito pequeno, e veja a diferença.

b) $x_{tot}^* = \sum_{i=1}^6 N_i \bar{x}_i$.

Como $\bar{x}_1 = 0$; $\bar{x}_2 = 7/26$, $\bar{x}_3 = 35/18$, ..., substituindo vem $x_{tot}^* = 40601.97$ ha é a estimativa da área total de trigo na região.

A precisão é dada pela variância $Var[X_{tot}^*]$ ou pelo desvio padrão $\sqrt{Var[X_{tot}^*]}$

A variância estimada é $\widehat{Var}[X_{tot}^*] = \sum_{i=1}^6 N_i^2 (1 - f_i) \frac{s_i'^2}{n_i}$

Para o seu cálculo necessitamos apenas de calcular $s_i'^2 = \frac{n_i \sum_{j=1}^{n_i} x_{ij}^2 - \left(\sum_{j=1}^{n_i} x_{ij}\right)^2}{n_i(n_i - 1)}$,

mas não temos os valores de $\sum_{j=1}^{n_i} x_{ij}^2$, para cada um dos estratos.

c) A estimativa é $\widehat{p}_{st} = \sum_{i=1}^6 W_i \hat{p}_i = 0.4341$

Observe-se que N_i é grande e podemos considerar a afectação proporcional pois $f = n/N = 0.0501 \approx n_i/N_i$ ($i = 1, \dots, 6$)

Então $Var[\widehat{P}_{st}] \simeq \frac{1-f}{n} \sum W_i P_i (1 - P_i)$, para calcular uma estimativa da variância, uma vez que não conhecemos as verdadeiras proporções, usamos \hat{p}_i , então

$$\widehat{Var}[\widehat{P}_{st}] \simeq \frac{1-f}{n} \sum W_i \hat{p}_i (1 - \hat{p}_i) = \frac{1-0.5}{125} \left(\frac{435}{2496} \times \frac{0}{22} + \frac{519}{2496} \times \frac{1}{26} + \dots \right) = \dots, \text{ é só fazer as contas...}$$

O intervalo de confiança a 95% será dado por $0.4341 \pm 1.96 \sqrt{\widehat{Var}[\widehat{P}_{st}]}$

d) Agora trata-se de uma estimativa de uma razão $r^* = \frac{\sum x_i}{\sum y_i}$, no último estrato.

$r^* = 79.36$ ha, i.e., esta é uma estimativa da área média de cada herdade com trigo.

e) Como $Y_T = 180$ esta informação permite usar o obter uma estimativa do total de X , usando o estimador da razão: $X_T^* = r_6^* Y_T = 14285.45445$ ha

f) Para comparar as estimativas é necessário determinar a variância de cada um dos estimadores $Var[\bar{X}_6]$ e $Var[\bar{X}_{6,R}]$

Não temos possibilidade de calcular como vimos atrás, mas ainda assim a estimativa da razão é preferível sempre que o coeficiente de correlação entre x e y , $\rho > \frac{1}{2} \frac{CV_Y}{CV_X}$

g) Nesta alínea vamos considerar vamos considerar que pretendem os recolher uma amostra total a recolher com a mesma dimensão $n = 125$.

Usando as fórmulas dadas nos slides finais tem-se para dimensão de amostra a recolher em cada estrato $n_i = \frac{n W_i \sqrt{P_i Q_i}}{\sum W_i \sqrt{P_i Q_i}}$.

Podemos simplificar um pouco e temos que usar as estimativas das proporções, donde $n_i = \frac{n N_i \sqrt{\hat{p}_i \hat{q}_i}}{\sum N_i \sqrt{\hat{p}_i \hat{q}_i}}$, ... agora é só umas contitas!!!

11. Complete o quadro calculando W_i e f_i

Estrato	N_i	n_i	\bar{x}_i	$s_i'^2$	W_i	f_i
A	5703.9	14	0.50	0.068		
B	1270.0	16	1.25	0.042		
C	1286.4	13	4.00	1.146		
D	5063.9	15	1.80	0.794		

As unidades de amostragem consideradas foram parcelas de $5 m^2$ de área.

- a) A média do nº de amêijoas/parcela é $\bar{x}_{st} = \sum_i \frac{N_i}{N} \bar{x}_i = 1.403$
 uma estimativa do total é $x_T^* = 18700.07$.
 A precisão da estimativa é dada por $s'^2(x_T^*) = 1662984$ ou pelo desvio padrão $s'(x_T^*) = 1289.567$
- b) I.C. a 95% para o n.º médio de amêijoas em cada parcela de $5 m^2$.
 $1.2132 < \mu < 1.5927$
- c) Dimensão total de amostra a recolher (e a dimensão em cada estrato)
 95% de confiança e $V = 0.002$
 Resposta: $n_1 = 31$ $n_2 = 6$ $n_3 = 29$ e $n_4 = 96$.

ANEXO

#topicos de resolução dos exercicios de amostragem

```
rm(list=ls())
#####
##exercicio 3
#####
##a)
###

x<-c(61,38,75,43,37,23,31,46,63,72,38,29,31,52,27,52,44,74,38,42)
y<-c(12,17,14,8,15,13,9,22,25,12,10,27,14,11,8,3,11,12,10,9)
N<-800
n<-20
```

```

sum(x)
sum(x^2)

#total Xt

Xt<-N*mean(x);Xt

##b)

z<-x+y;z
sum(z)
sum(z^2)
mean(z)*800
sd(z)
var(z)
#estimativa da variância do estimador do total

s2t<-N^2*(1-20/800)*var(z)/n;s2t
##c)

#####I:C para o total

800*mean(z)-qt(0.995,19)*800*sd(z)*sqrt((1-20/800)/20)
800*mean(z)+qt(0.995,19)*sd(z)*sqrt((1-20/800)/20)*800

####
#d)
#####
n0<-4*var(x)/(3*3);n0

n<-n0/(1+n0/800);n

#####
##e)
#####
r<-sum(x)/sum(y);r
f<-20/800
var_r<-(1-f)*(sum(x^2)-2*r*sum(x*y)+r^2*sum(y^2))/(20*19*mean(y)*mean(y))
var_r

rm(list=ls())
#####

```



```

##exercicio 5
#####

# sendo um rectângulo de 5 km x 4km, o nº de parcelas será 50x40=2000 parcelas
#i=1....40,
#j=1...50

n<-20
linhas<-seq(1:40);colunas<-seq(1:50)
i<-sample(linhas,n);i
j<-sample(colunas,n);j
amostra<-cbind(i,j);amostra

#####
##exercicio 6
#####

## a)

x1<-c(9.0,8.2, 7.8, 8.0,6.7, 12.3, 7.7, 8.3,
7.8, 8.7,9.4,10.1, 9.5, 6.4, 9.2, 6.6)
x1
sum(x1)
sum(x1*x1)
mean(x1)
var(x1)
var(x1)/16
sqrt(var(x1)/16)

x2<-c(3.6 , 5.3 , 7.0, 6.3, 9.2, 5.0, 4.0 , 4.1,
3.7, 9.1, 2.4, 5.3, 6.4, 4.3, 3.9, 6.1,
3.8 , 3.0 , 6.5 , 3.2, 4.3, 7.3, 3.0, 6.3,
3.2, 7.2, 6.0, 6.2, 6.9,5.1, 4.7,6.1,
2.8 , 8.4, 2.9, 4.1, 3.9, 6.3,10.0,5.2)
x2
length(x2)
sum(x2)
sum(x2*x2)
mean(x2)
var(x2)
var(x2)/40
sqrt(var(x2)/40)

```

```

xtot<-c(x1,x2)
sum(xtot)
mean(xtot)
var(xtot)
sqrt(var(xtot)/56)

mst<-16/56*mean(x1)+40/56*mean(x2)

(16/56)^2*var(x1)/16+(40/56)^2*var(x2)/40

sd<-sqrt((16/56)^2*var(x1)/16+(40/56)^2*var(x2)/40)

##b)

mst-1.96*sd; mst+1.96*sd

## c)

base<-16/56*sd(x1)+40/56*sd(x2);base

n1<-16/56*sd(x1)/base*60;n1
n2<-40/56*sd(x2)/base*60;n2

rm(list=ls())
#####
##exercicio 7
#####
N<-588
n<-25

small<-c(17, 38, 9, 7, 11, 14, 17, 10, 31, 24, 22, 21, 9, 41, 19,
  9, 13, 26, 36, 18, 8, 11, 23, 19, 16, 14, 14, 17, 20, 20,
  9, 18, 6, 19, 52, 14, 5, 27, 14, 14, 28, 17, 9, 11, 12,
  25, 19, 28, 15, 18, 24, 23, 27, 24, 20, 21, 27, 21, 34, 26,
  21, 9, 29, 22, 10, 18, 45, 24, 16, 95, 40, 42, 11, 17, 17,
  13, 14, 23, 17, 27, 18, 34, 18, 16, 17, 20, 23, 18, 42, 22,
  18, 23, 16, 26, 11, 37, 23, 32, 24, 16, 24, 34, 37, 31, 29,
  15, 41, 38, 21, 34, 23, 24, 27, 34, 5, 34, 29, 22, 26, 30,
  26, 27, 39, 30, 31, 28, 39, 28, 34, 28, 24, 44, 22, 23, 40,
  16, 5, 19, 36, 36, 17, 21, 43, 21, 19, 14, 14, 31, 27, 39,
  30, 41, 28, 19, 32, 18, 19, 33, 27, 28, 26, 23, 32, 36, 21,

```

```
24, 32, 19, 18, 31, 25, 26, 21, 18, 36, 29, 47, 26, 31, 26,
32, 27, 43, 45, 45, 25, 17, 30, 27, 28, 16, 44, 20, 15, 31,
21, 42, 27, 32, 33, 21, 35, 44, 24, 26, 38, 57, 54, 24, 37,
21, 33, 19, 20, 32);small
```

```
mean(small)
```

```
medio<-c(37, 30, 41, 17, 38, 29, 32, 21, 39, 41, 28, 33, 35, 24, 36,
28, 20, 23, 27, 34, 33, 36, 25, 28, 39, 36, 22, 25, 54, 53,
36, 14, 22, 32, 21, 35, 35, 39, 32, 40, 24, 48, 41, 30, 42,
20, 38, 23, 17, 38, 16, 23, 28, 32, 18, 60, 28, 47, 61, 25,
22, 25, 48, 53, 35, 25, 23, 44, 18, 56, 42, 55, 39, 24, 38,
42, 27, 30, 34, 43, 29, 35, 43, 62, 25, 15, 66, 34, 25, 11,
45, 28, 40, 32, 38, 33, 48, 46, 54, 45, 35, 31, 30, 42, 22,
23, 46, 14, 42, 33, 31, 75, 50, 44, 33, 41, 32, 45, 44, 51,
39, 35, 22, 44, 35, 24, 29, 23, 32, 30, 35, 50, 28, 21, 21,
12, 30, 28, 60, 35, 49, 33, 22, 58, 25, 23, 39, 40, 44, 41,
14, 37, 32, 22, 27, 23, 37, 59, 50, 46, 40, 47, 41, 38, 48,
40, 32, 31, 22, 24, 25, 33, 54, 36, 52, 39, 61, 46, 36, 16,
37, 38, 51, 25, 35, 49, 9, 46, 35, 53, 43, 59, 41, 52, 51,
47, 72, 46, 29, 25, 42, 42, 43, 46, 43, 29, 58, 47, 85, 52,
48, 23, 39, 40, 43, 52, 36, 35, 27, 56, 47, 39, 51, 48, 48,
23, 24, 39, 30, 59, 35, 39, 32, 51, 18, 27, 38, 36, 41, 11,
42, 42, 65, 27, 34, 72, 49, 39, 44, 57, 64, 51, 53, 55, 63,
39, 31, 48);medio
```

```
large<-c(53, 63, 44, 66, 40, 42, 48, 44, 27, 56, 37, 39, 37, 40, 66,
49, 39, 54, 30, 68, 36, 42, 28, 29, 41, 57, 30, 39, 28, 80,
79, 61, 81, 53, 57, 54, 29, 94, 77, 52, 61, 49, 52, 67, 36,
35, 57, 63, 32, 48, 57, 50, 62, 51, 52, 59, 55, 22, 18, 84,
57, 86, 50, 54, 96, 45, 28, 59, 64, 42, 41, 77, 76, 83, 36,
42, 39, 72, 84, 34, 55, 51, 66, 96, 63, 88, 87, 63, 91, 117,
107, 48, 56, 71, 54, 64, 45, 61, 59, 68, 50, 74, 100, 144, 80,
64, 101, 105, 77, 85, 60, 63, 66, 36, 95);large
```

```
xmed<-mean(small);xmed
length(small)
```

```
# se quisermos só trabalhar com as "small" fazemos
```

```

pos<-sample(length(small),25,rep=F);pos
x1<-c(small[c(pos)]);x1 #temos então já a nossa amostra

#Se quisermos obter sempre a mesma amostra temos que fazer set.seed( )
set.seed(5)#modo de estrairmos sempre a mesma
x2<-sample(small,25,rep=F);x2
set.seed(5)
pos<-sample(length(small),25,rep=F);pos
x1<-c(small[c(pos)]);x1 #temos então já a nossa amostra

##### MAS VAMOS AO EXERCICIO
#####
#a)
#####

#Agora pede-se que se retire 25 de toda a população

pop<-c(small,medio,large);pop
mean(pop);var(pop)
sum(pop)

length(pop)
x3<-sample(pop,25,rep=F);x3

#####
#b)
#####

mean(x3)
xt<-length(pop)*mean(x3);xt #temos uma estimativa de Total

#Como conhecemos a população podemos usar o verdadeiro valor da variância

#I.C a 95% para o valor médio

f<-25/length(pop)
mean(x3)-qnorm(0.975)*sd(pop)*sqrt((1-f)/25)
mean(x3)+qnorm(0.975)*sd(pop)*sqrt((1-f)/25)

```

```

#para o total basta multiplicar por 588

#####
#d)
#####
N1<-length(small)
N2<-length(medio)
N3<-length(large)
N1+N2+N3

n1<-9
n2<-11
n3<-5

xs<-sample(small,n1,rep=F);xs
xm<-sample(medio,n2,rep=F);xm
xl<-sample(large,n3,rep=F);xl
mean(xs);mean(xm);mean(xl)
var(xs);var(xm);var(xl)

#####
#e)
#####

xmean_str<-(N1*mean(xs)+N2*mean(xm)+N3*mean(xl))/N
xmean_str

rm(list=ls())
###
#exercicio 8

##a)

n.pessoas<-c(5, 6, 3, 3, 2, 3, 4, 4, 3,
3, 4 , 3, 3, 1, 2, 4, 3, 4, 2, 4)
n<-20
N<-12400

x.bar<-mean(n.pessoas);x.bar
e.total<-N*x.bar;e.total

```

```

###
##b)I. Conf.

l.inf<-e.total-qt(0.05,n-1,lower.tail=F)*N*sd(n.pessoas)*sqrt((1-n/N)/n);l.inf
l.sup<-e.total+qt(0.05,n-1,lower.tail=F)*N*sd(n.pessoas)*sqrt((1-n/N)/n);l.sup

#ou

l.inf<-e.total-qt(0.975,n-1)*N*sd(n.pessoas)*sqrt((1-n/N)/n);l.inf
l.sup<-e.total+qt(0.975,n-1)*N*sd(n.pessoas)*sqrt((1-n/N)/n);l.sup

n.aut<-c(2, 3, 1, 0, 1, 2, 0, 3, 1, 0, 1, 0, 2, 1, 0, 2,
1, 0, 1, 3)
r<-sum(n.pessoas)/sum(n.aut);r
N.aut<-11000
t.pessoas<-r*N.aut;t.pessoas

ybar<-mean(n.aut);ybar
xmean.bar<-r*11000/12400

rm(list=ls())
###
##exercicio 10
###
Ni<-c(5703.9,1270,1286.4,5063.9)
ni<-c(14,16,13,15)
xbar.i<-c(0.5,1.25,4,1.8)
fi<-ni/Ni;fi
var.i<-c(0.068,0.042,1.146,0.796)
N<-sum(Ni);N

#####
#a)
#####
tot.est<-sum(Ni*xbar.i);tot.est
var.tot<-sum(Ni*Ni*(1-fi)*var.i/ni);var.tot

var.mean<-var.tot/(N*N);var.mean
sd.mean<-sqrt(var.mean);sd.mean

#####

```

```
#d)
#####

yi<-c(1,2,5,2)
pi<-yi/ni
pi
#[1] 0.07142857 0.12500000 0.38461538 0.13333333
pst<-sum(Ni*pi)/N;pst
#[1] 0.1302988
```