

Disciplina: AMOSTRAGEM

Prof.: Pedro Luis do Nascimento Silva

Aluno: Marcel Dantas de Quintela

Mat. : 201230131-27

Data: 19.12.2012

## **Relatório Exame Final**

### **1 – INTRODUÇÃO**

O presente trabalho tem como objetivo avaliar comparativamente os cinco planos amostrais dos quais se incluem o plano por Amostra Aleatórios Simples e planos Estratificados por “alocação igual” e “por corte”.

Serão utilizados para estudar o desempenho esperado os dados do arquivo Munic1.xlsx, composto pelas seguintes informações dos 5565 municípios: Código do Município, Nome do Município, Código da UF, Sigla UF, População estimada do município em 1 de agosto de 2011, Total de funcionários ativos da administração direta, Existe Maternidade no município, Existe Unidade de emergência no município. Destas, as últimas três são nossas “variáveis de estudo”.

Diante dessas informações, deverá ser eleito o melhor plano o qual será usado para selecionar uma amostra com 200 municípios capaz de estimar os seguintes parâmetros populacionais:

- Total de funcionários ativos da administração direta (Total.FuncADMD);
- Razão da população por funcionário ativo da administração direta (Razão.Pop/Func);
- Proporção de municípios com maternidade (Prop.Mater); e
- Proporção de municípios com maternidade e emergência (Prop.Mat&Eme).

### **2 – METODOLOGIA**

Devido a falhas no cadastro da variável de interesse FuncADMD, foram removidos do universo os municípios de: Abaré-BA (recusa), Palmas-TO (ignorado), Campina Grande-PB (ignorado), Mirandiba-PE (ignorado), Macaé-RJ (ignorado) e Fernando de Noronha-PE (variável zerada).

Após realizada a remoção destes municípios o tamanho das unidades amostrais reduziu para 5.559 municípios.

## Amostra Aleatória Simples – AAS

É o procedimento de seleção em que as amostras possíveis de tamanho  $n$ , obtidas em um processo sem reposição da unidade amostral, tenham a mesma probabilidade de serem escolhidas.

- Variância para o Estimador de Total

$$V_{AAS}(\hat{Y}_{HT}) = N^2 \left( \frac{1}{n} - \frac{1}{N} \right) S_y^2, \text{ onde: } S_y^2 = \frac{1}{N-1} \sum_{i \in U} (y_i - \bar{Y})^2$$

- Variância para o Estimador de Razão

Devido ao um pequeno vício técnico<sup>1</sup> no estimador  $\hat{R}$ , ao invés de avaliarmos sua variância usaremos seu Erro Quadrático Médio (EQM).

$$V_{AAS}(\hat{R}) = EQM_{AAS}(\hat{R}) = \frac{(1-f)}{n\bar{X}^2} * \frac{\sum_{i \in U} (y_i - R x_i)^2}{N-1};$$

$$\text{Onde: } \hat{R} = \frac{\hat{Y}_{HT}}{\hat{X}_{HT}} \rightarrow R_{AAS} = \frac{\sum_{i \in S} y_i}{\sum_{i \in S} x_i} = \frac{\bar{y}}{\bar{x}} \text{ e } f = n/N$$

- Variância para o Estimador de Proporção

$$V_{AAS}(p) \approx \left( \frac{N-n}{N-1} \right) \frac{P(1-P)}{n}, \text{ se } \frac{n}{N} \text{ é pequena;}$$

$$\text{Onde: } P = \frac{T}{N} \text{ (média populacional) e } T = \sum_{i \in U} y_i$$

## Amostra Estratificada Simples – AES

Processo em que a população é dividida em subgrupos mutuamente exclusivos e exaustivos, visando, principalmente, aumentar a precisão das estimativas. As amostras são selecionadas dentro de cada estrato. Em cada um deles são estimados os parâmetros de interesse conforme processo de seleção definido. As estimativas são agregadas para o conjunto da população.

Sendo o universo  $U$  dividido em  $H$  grupos de forma que:  $U = U_1 \cup U_2 \cup \dots \cup U_H = \bigcup_{h=1}^H U_h$  e  $U_h \cap U_k = \emptyset, h \neq k$ . Então:  $U_h = [i: \text{unidade amostral } i \in \text{ao estrato } h], h = 1, 2, \dots, H$ . Desta forma, sendo  $N_h$  o tamanho de  $U_h$ , então:  $N_1 + N_2 + \dots + N_H = N$ .

Sendo selecionada uma amostra  $s_h$  de tamanho  $n_h$ , com  $n_h > 0$ , segundo um plano amostral independentemente em cada estrato  $h$ , onde  $\sum_{h=1}^H n_h = n$ , assegura-se que cada estrato tem sua população completa por:  $s = s_1 \cup s_2 \cup \dots \cup s_H$ .

No caso da AES, a AAS é empregada em todos os estratos tendo a variância definida nos seguintes estimadores:

- Variância para o Estimador de Total

$$V_{AES}(\hat{Y}_{AES}) = N^2 \sum_h W_h^2 \left( \frac{1}{n_h} - \frac{1}{N_h} \right) S_h^2. \text{ Onde: } W_h = \frac{N_h}{N} \text{ e } S_h^2 = \frac{1}{N_h-1} \sum_{i=1}^{N_h} (y_{hi} - \bar{Y}_h)^2$$

---

<sup>1</sup> Vício na ordem de  $n^{-1}$  e fica menor à medida que o tamanho da amostra cresce, geralmente é desprezado na prática, a não ser que as amostras sejam muito pequenas.

- Variância para o Estimador de Razão

$$V_{AES}(\hat{Y}_{Rc}) = \sum_h \frac{N_h^2(1-f_h)}{n_h} (S_{yh}^2 + R^2 S_{xh}^2 - 2R\varphi_h S_{yh} S_{xh})$$

$$\text{Onde: } f = \frac{n_h}{N_h}, \hat{R} = \frac{\sum_{h=1}^H \bar{y}_h}{\sum_{h=1}^H \bar{x}_h}, S_{yh}^2 = \frac{1}{N_h-1} \sum_{i=1}^{N_h} (y_{hi} - \bar{Y}_h)^2, S_{xh}^2 = \frac{1}{N_h-1} \sum_{i=1}^{N_h} (x_{hi} - \bar{X}_h)^2$$

- Variância para Estimador de Proporção

$$V_{AES}(p_{AES}) = \sum_h W_h^2 \left( \frac{1}{n_h} - \frac{1}{N_h} \right) S_h^2. \text{ Onde: } W_h = \frac{N_h}{N} \text{ e } S_h^2 = \frac{1}{N_h-1} \sum_{i=1}^{N_h} (y_{hi} - \bar{Y}_h)^2$$

## Planos

1. Amostra Aleatória Simples – AAS

Correspondente à seleção de municípios referenciada a metodologia acima citada.

2. Amostra Estratificada Simples – AES1

Neste plano foram utilizados os estratos naturais, Grandes Regiões do Brasil, com alocação igual da amostra, onde  $n = 200$ , foi dividido igualmente entre as 5 Regiões, tendo  $n_i = 40$ .

3. Amostra Estratificada Simples – AES2

Semelhante ao Plano 2, sua alteração foi na forma de definição dos estratos. Usando de estratificação estatística onde cada um dos estratos tivesse a soma da raiz quadrada da população dos municípios fosse aproximadamente igual.

4. Amostra Estratificada por Corte – AEC

Para este plano foram definidos dois estratos, sendo o primeiro (estrato certo), contendo todos os municípios que são capitais e os que têm população maior ou igual a 500.000 habitantes. No segundo estrato será aplicada AAS. Sua alocação ficou definida como  $N_1 = n_1 = 44$  e  $N_2 = 5515; n_2 = 156$ .

Recordamos que a capital do Tocantins, Palmas, foi removida das possíveis amostras devido a falhas cadastrais.

5. Amostra Aleatória Simples – AAS2

Esse plano usará a Amostragem aleatória simples de municípios (igual ao plano 1) utilizando o estimador do tipo razão tendo como variável auxiliar o total da população. Onde:

Variância para o estimador de Total

$$V(\hat{Y}) = \frac{N^2(1-f)}{n} \left[ \sum_{i \in U} \frac{(y_i - Rx_i)^2}{N-1} \right]$$

$$\hat{Y} = \frac{\bar{y}}{\bar{x}} X, X \text{ é o total populacional}$$

Variância para o estimador de Proporção

$$V(\hat{Y}) = \frac{(1-f)}{n} \left[ \sum_{i \in U} \frac{(y_i - Rx_i)^2}{N-1} \right]$$

$$\hat{Y} = \frac{\bar{y}}{\bar{x}} \bar{X}, \bar{X} \text{ é a média populacional}$$

Para este plano não está definida variância para a razão entre a população e o número de funcionários ativos da administração direta.

## Cr terios para sele  o do Plano Amostral

Para avaliarmos o desempenho de cada plano amostral, foram utilizadas como medidas para cada estimador, o Coeficientes de Varia  o (CV) e o Efeito do Plano Amostral(EPA), denotados abaixo:

$$CV_j(Y) = \frac{\sqrt{V_j(Y)}}{Y_j} * 100$$

Onde  $j$  s o os planos estudados.

$$EPA_j(Y) = \frac{V_j(Y)}{V_{AAS}(Y)} * 100$$

O plano de melhor desempenho ser  o que apresentar, conjuntamente, menores CV's e EPA's para seus estimadores.

## Sele  o da amostra

A sele  o das amostras ser  realizada atrav s de execu  o da programac  o feita no software R com o apoio do pacote *samplin* (Ver ANEXO 1).

## C lculo das estimativas

As estimativas das vari ncias  $\hat{V}_{plano}(\hat{Y})$  ser o calculadas com base em seu estimador  $V_{plano}(\hat{Y})$ , alterando em cada f rmula  $\sum_{i \in U}$  para  $\sum_{i \in s}$  nos planos AAS e  $\sum_{h=1}^H$  para  $\sum_{s=1}^S$  nos planos AES e AEC.

## 3 – RESULTADOS

Para compararmos o desempenho de cada plano amostral conforme os cr terios, definimos na Tabela 1 os resultados para os Coeficientes de Varia  o,

**Tabela 1: Coeficientes de Varia  o para os Estimadores dos Planos Amostrais.**

| Planos | Coeficiente de Varia  o |                |               |               |
|--------|-------------------------|----------------|---------------|---------------|
|        | Total.FuncADMD          | Raz o.Pop/Func | Prop.Mater    | Prop.Mat&Eme  |
| 1      | 22,3704                 | 22,0520        | 8,4423        | 9,5361        |
| 2      | 24,8254                 | 24,1182        | 9,5270        | 10,7449       |
| 3      | 9,7402                  | 7,8425         | <b>8,2326</b> | <b>8,9308</b> |
| 4      | <b>8,7053</b>           | <b>4,7998</b>  | 9,5104        | 10,7323       |
| 5      | 22,0520                 | NA             | 41,2863       | 41,3151       |

Temos que os planos 3 e 4 apresentaram melhores resultados, sendo os estimadores de Total de Funcion rios da Administra  o Direta e o de Raz o entre Popula  o e Funcion rios com melhor desempenho no Plano 4; e os estimadores de Propor  o de Maternidades e o de Propor  o de Maternidade e Emerg ncia menores varia  es conforme o Plano 3.

Dado a indefini  o entre os dois planos, apresentamos na Tabela 2 os resultados para o Efeito do Plano Amostral tomado como base o Plano 1, como forma de refinamento no cr terio de avalia  o entre os planos considerados melhores para representar os estimadores.

**Tabela 2: Efeito Plano Amostral para os Estimadores dos Planos Amostrais**

| Planos | EPA (Base Plano1-AAS) |                |            |              |
|--------|-----------------------|----------------|------------|--------------|
|        | Total.FuncADMD        | Razão.Pop/Func | Prop.Mater | Prop.Mat&Eme |
| 2      | 1,1097                | 1,0937         | 1,1285     | 1,1267       |
| 3      | 0,4354                | 0,3556         | 0,9752     | 0,9365       |
| 4      | 0,3891                | 0,2177         | 1,1265     | 1,1254       |
| 5      | 0,9858                | NA             | 4,8904     | 4,3325       |

O Plano 2 perdeu eficiência em relação ao Plano 1, essa perda sugere que os estratos regionais não sejam uma boa forma de estratificação para os estimadores estudados.

O Plano 4 mostrou ganhos nos dois estimadores constituídos por variáveis contínuas, como a variabilidade no estrato certo foi anulada, as unidades amostrais que compuseram o Estrato 2 foram responsáveis pela variância do Plano. Removidos os municípios mais representativos a variabilidade para os estimadores que dependem das variáveis Total.FuncADMD e Razão.Pop/Func teve ganho bem maior que em outros planos. Contudo, os estimadores compostos pela variável indicadora manteve praticamente o mesmo comportamento da do Plano2.

O Plano 3 apresentou melhor desempenho em relação ao Plano 1 (AAS), que os demais planos estudados existindo ganho no desempenho em todos os estimadores em relação aos demais planos. Sendo o escolhido para selecionar a amostra da pesquisa solicitada.

Para o processo e seleção da amostra utilizou-se dos recursos do software estatístico *R* a função *strata* do pacote *sampling*, a qual por meio do atributo *srswor*, definiu que amostragem seria aleatória simples e sem reposição. Selecionada a amostra, procedemos com sua modelagem através da função *svydesign{survey}* que nos permitiu estimar o Total.FuncADMD (*svytotal*), Razão.Pop/Func (*svyratio*) e as Proporções Prop.Mater e Prop.Mat&Eme (*svymean*). Os resultados desse processo foram resumidos na Tabela 3.

Os detalhes da programação em ANEXO e disponibilizados pelo autor em: [https://github.com/Marcelquintela/Amost/blob/master/code\\_r](https://github.com/Marcelquintela/Amost/blob/master/code_r).

**Tabela 3: Resumo Estatístico da Amostra de 200 Municípios Selecionada conforme o Plano 3**

| Variável       | Parâmetro | Estimativa | Erro padrão | CV       |
|----------------|-----------|------------|-------------|----------|
| Total.FuncADMD | 5.637.625 | 5.986.974  | 684.544     | 11,43389 |
| Razão.Pop/Func | 33,9699   | 37,0292    | 3,5302      | 9,5337   |
| Prop.Mater     | 0.4035    | 0.4175     | 0.0336      | 8,0402   |
| Prop.Mat&Eme   | 0.3465    | 0.3567     | 0.0311      | 8,7284   |

Considerando a amostra selecionada, é possível notar que as estimativas pontuais são semelhantes ao parâmetro populacional de interesse. Todas as estimativas pertencem ao intervalo de confiança em pelo menos 95%. Sendo assim, o Plano 3 consegue representar muito bem a população quanto as variáveis de interesse desse estudo.

#### 4 – DISCUSSÕES

Levando em conta que o plano amostral selecionado como melhor entre os propostos para este trabalho foi proposto sob Amostragem Estratificada, fatores que influenciam na eficiência de planos amostrais dessa natureza são: Escolha da(s) variável(is) de estratificação; o Número de estratos; Os limites dos estratos; a forma de alocação nos estratos; e o método de seleção em cada estrato.

Como proposta de melhoria do Plano 3. Buscaria com uso dos recursos disponibilizado pela biblioteca *strata.LH {stratification}* do software *R*, ajustar por meio da Alocação Ótima os Limites e o Número(máximo 6 conforme – Cochran(1997)) de estratos, verificando qual configuração seria a de menor variância e que proporcionasse menor custo.

## ANEXO

```
## *****##
# Exame final Amostragem 2012
# Aluno Marcel Dantas de Quintela Mat. 2012303131-27
## *****##

setwd("D:\\Ence_Mestrado_2012\\03.Amostragem\\Exercícios\\Exame_Final") # Define pasta de trabalho
getwd() # Verifica se alocação da pasta de trabalho foi correta
ls() # Lista arquivos encontrados na pasta de trabalho

# Carrega bibliotecas requeridas para o trabalho
library(foreign) # Carrega biblioteca para ler arquivo de dados em formato 'externo'
library(stratification)
library(sampling) # carrega pacote para seleção de amostra
library(survey) # carrega pacote de análise amostral

munic = read.csv("Munic.csv", sep=";") # Leitura dos dados da população de fazendas
munic = transform(munic,
  FuncADMD = as.numeric(FuncADMD),
  Populacao = as.numeric(Populacao),
  Maternidade=ifelse(Maternidade=="Sim",1,ifelse(Maternidade=="Não",0,NA)),
  Emergencia=ifelse(Emergencia=="Sim",1,ifelse(Emergencia=="Não",0,NA)),
  Razao_Pop.Func = (Populacao/FuncADMD),
  Mat_Emer = ifelse(Maternidade=="Sim" & Emergencia=="Sim",1,0))

# Criando a variável capitais, necessária para o Plano4
# Lembrando que Palmas foi removida do cadastro e o geocódigo foi removido o último dígito
capitais <- c('110020','120040','130260','140010','150140','160030','172100','211130','221100',
  '230440','240810','250750','261160','270430','280030','292740','310620','320530',
  '330455','355030','410690','420540','431490','500270','510340','520870','530010')
munic = transform(munic,Capitais = ifelse(is.element(munic$CodMunic,capitais),1,0))
#rm(munic)

(N = nrow(munic)) # Total de municípios
(n = 200) # tamanho da amostra
(f = n/N) # Fração Amostral

raz.soma = function(n, Estrato) { #Razão das somas para os estratos
  a = rep(0,n)
  for (i in 1:n){a[i]=sum(munic$Populacao[Estrato==i])/sum(munic$FuncADMD[Estrato==i])}
  return(a)
}

Corr = function(n,Estrato){
  a = numeric(n)
  for(i in 1:n) {
    a[i] = cor(munic$Populacao[Estrato==i],munic$FuncADMD[Estrato==i])
  }
  return(a)
}

# Var.prop = function(N,n,P){
# v = ((N-n)/(N-1)) * P * (1-P)*(1/n)
# }

Resultados = matrix(c(rep(NA,20)),nrow=5, ncol=4, byrow=T,
  dimnames=list(c("Plano 1", "Plano 2", "Plano 3", "Plano 4", "Plano 5"),
    c("Total.FuncADMD", "Razão.Pop/Func", "Prop.Mater", "Prop.Mat&Eme" )))

##### Plano 1 - Amostragem Aleatória Simples de municípios
# a) Total de funcionários ativos da administração direta
tot.a = sum(munic$FuncADMD, rm.na=T)
(Plano1.Vartot.a = (N^2) * (1/n - 1/N) * var(munic$FuncADMD))
(Plano1.CVtot.a = 100*sqrt(Plano1.Vartot.a) / tot.a)
```

```

# b) Razão da população por funcionário ativo da administração direta ?????
(tot.b = sum(munic$Populacao)/sum(munic$FuncADMD))
(med_raz = mean(munic$FuncADMD))
(Plano1.Vartot.b = (1-f)/(n*med_raz^2)*1/(N-1)*sum((munic$Populacao-tot.b*munic$FuncADMD)^2))
(Plano1.CVtot.b = 100*sqrt(Plano1.Vartot.b)/tot.b)

# c) Proporção de municípios com maternidade
prop.c = sum(munic$Maternidade)/N
(Plano1.Vartot.c = ((N-n)/(N-1))*prop.c*(1-prop.c)/n)
(Plano1.CVtot.c = 100*sqrt(Plano1.Vartot.c)/prop.c)

# d) Proporção de municípios com maternidade e emergência
prop.d = sum(munic$Mat_Emer)/N
(Plano1.Vartot.d = ((N-n)/(N-1))*prop.d*(1-prop.d)/n)
(Plano1.CVtot.d = 100*sqrt(Plano1.Vartot.d)/prop.d)

Totais= c(tot.a,tot.b,prop.c,prop.d)

Resultados[1,] = c(Plano1.CVtot.a,Plano1.CVtot.b,Plano1.CVtot.c,Plano1.CVtot.d)

##### Plano 2 - Amostragem estratificada simples de municípios, com os municípios
####          estratificados por região e alocação igual
#cria estratos para o Plano 2
munic = transform(munic,Estrato2=substr(CodMunic,1,1))

###Função para geral a razão das Somas para cada estrato
Estrato2.Nh = as.vector(table(munic$Estrato2))
Estrato2.nh = rep(200/5,5)
Wh=Estrato2.Nh/N

Estrato2.Yh = aggregate(cbind(FuncADMD,Razao_Pop.Func)~Estrato2, data=munic,FUN=sum)
aux = aggregate(cbind(FuncADMD,Razao_Pop.Func,Maternidade,Mat_Emer)~Estrato2, data=munic,FUN=mean)
aux[,3]= raz.soma(5, munic$Estrato2)
Estrato2.Yh = cbind(Estrato2.Yh,aux[,-1])

Estrato2.Sh2 = aggregate(cbind(FuncADMD,Razao_Pop.Func,Maternidade,Mat_Emer,Populacao)~Estrato2,
                        data=munic,FUN=var)
Estrato2.SD = aggregate(cbind(FuncADMD,Razao_Pop.Func,Maternidade,Mat_Emer,Populacao)~Estrato2,
                        data=munic,FUN=sd)

Vartot.raz = (1/sum(munic$FuncADMD)^2)*sum(Estrato2.Nh^2 * (1- (Estrato2.nh/Estrato2.Nh))/Estrato2.nh *
      (Estrato2.Sh2[6] + Estrato2.Yh[5]^2*Estrato2.Sh2[2]
      - 2 * Estrato2.Yh[5] * Estrato2.SD[6] * Estrato2.SD[2]
      * Corr(5,munic$Estrato2)))
Estrato2.Sh2 = Estrato2.Sh2[,c(1,6)]

(Plano2.Vartot = colSums(Wh^2 * ((1 / Estrato2.nh) - (1 / Estrato2.Nh)) * Estrato2.Sh2))
Plano2.Vartot[1] = N^2*Plano2.Vartot[1] ## Correção para o estimador de Total de Funcionários
Plano2.Vartot[2] = Vartot.raz
(Plano2.CVtot = 100*sqrt(Plano2.Vartot) / Totais)

rm(aux,Estrato2.Nh,Estrato2.nh,Estrato2.Yh,Estrato2.Sh2,Estrato2.SD )
Resultados[2,] = Plano2.CVtot

##### Plano 3 - Amostragem estratificada simples de municípios, estratificada por tamanho
#          (população) em cinco classes tais que a soma da raiz quadrada da população dos
#          municípios seja aproximadamente igual, e usando alocação igual da amostra nos estratos;
#cria estratos para o Plano 3
munic = transform(munic,sqrtpop=sqrt(Populacao))
munic = munic[order(munic$sqrtpop),]
aux = matrix(NA,nrow=nrow(munic),ncol=5)
for(i in 1:5){aux[,i] = cumsum(munic$sqrtpop) <= (sum(munic$sqrtpop)/5)*i}
munic = transform(munic,Estrato3=(6-rowSums(aux)))
munic = transform(munic,Estrato3=as.factor(Estrato3))

Estrato3.Nh = as.vector(table(munic$Estrato3))

```



```

Estrato3.nh = rep(200/5,5)
Wh=Estrato3.Nh/N

Estrato3.Yh = aggregate(cbind(FuncADMD,Razao_Pop.Func)~Estrato3, data=munic,FUN=sum)
aux = aggregate(cbind(FuncADMD,Razao_Pop.Func,Maternidade,Mat_Emer)~Estrato3, data=munic,FUN=mean)
aux[,3]= raz.soma(5, munic$Estrato3)
Estrato3.Yh = cbind(Estrato3.Yh,aux[,-1])

Estrato3.Sh2 = aggregate(cbind(FuncADMD,Razao_Pop.Func,Maternidade,Mat_Emer,Populacao)~Estrato3,
  data=munic,FUN=var)
Estrato3.SD = aggregate(cbind(FuncADMD,Razao_Pop.Func,Maternidade,Mat_Emer,Populacao)~Estrato3,
  data=munic,FUN=sd)

Vartot.raz = (1/sum(munic$FuncADMD)^2)*sum(Estrato3.Nh^2 * (1- (Estrato3.nh/Estrato3.Nh))/Estrato3.nh *
  (Estrato3.Sh2[6] + Estrato3.Yh[5]^2*Estrato3.Sh2[2]
  - 2 * Estrato3.Yh[5] * Estrato3.SD[6] * Estrato3.SD[2]
  * Corr(5,munic$Estrato3)))
Estrato3.Sh2 = Estrato3.Sh2[,-c(1,6)]

(Plano3.Vartot = colSums(Wh^2 * ((1 / Estrato3.nh) - (1 / Estrato3.Nh)) * Estrato3.Sh2))
Plano3.Vartot[1] = N^2*Plano3.Vartot[1] ## Correção para o estimador de Total de Funcionários
Plano3.Vartot[2] = Vartot.raz
(Plano3.CVtot = 100*sqrt(Plano3.Vartot) / Totais)

#rm(aux,Estrato3.Nh,Estrato3.nh,Estrato3.Yh,Estrato3.Sh2,Estrato3.SD )
Resultados[3,] = Plano3.CVtot

##### Plano 4 - Amostragem estratificada 'por corte', com inclusão de municípios com certeza
# na amostra. Neste plano, inclua todos os municípios de capitais e também todos os
# municípios com população igual ou superior a 500.000 habitantes com certeza na amostra.
# Selecione uma amostra aleatória simples dos municípios restantes.
munic = transform(munic,Estrato4 = ifelse(munic$Populacao>=500000|munic$Capitais==1, 1,2))
munic = transform(munic,Estrato4=as.factor(Estrato4))

Estrato4.Nh = as.vector(table(munic$Estrato4))
Estrato4.nh = c(Estrato4.Nh[1],200-Estrato4.Nh[1])
Wh=Estrato4.Nh/N

Estrato4.Yh = aggregate(cbind(FuncADMD,Razao_Pop.Func)~Estrato4, data=munic,FUN=sum)
aux = aggregate(cbind(FuncADMD,Razao_Pop.Func,Maternidade,Mat_Emer)~Estrato4, data=munic,FUN=mean)
aux[,3]= raz.soma(2, munic$Estrato4)
Estrato4.Yh = cbind(Estrato4.Yh,aux[,-1])

Estrato4.Sh2 = aggregate(cbind(FuncADMD,Razao_Pop.Func,Maternidade,Mat_Emer,Populacao)~Estrato4,
  data=munic,FUN=var)
Estrato4.SD = aggregate(cbind(FuncADMD,Razao_Pop.Func,Maternidade,Mat_Emer,Populacao)~Estrato4,
  data=munic,FUN=sd)

Vartot.raz = (1/sum(munic$FuncADMD)^2)*sum(Estrato4.Nh^2 * (1- (Estrato4.nh/Estrato4.Nh))/Estrato4.nh *
  (Estrato4.Sh2[6] + Estrato4.Yh[5]^2*Estrato4.Sh2[2]
  - 2 * Estrato4.Yh[5] * Estrato4.SD[6] * Estrato4.SD[2]
  * Corr(5,munic$Estrato4)))
Estrato4.Sh2 = Estrato4.Sh2[,-c(1,6)]

(Plano4.Vartot = colSums(Wh^2 * ((1 / Estrato4.nh) - (1 / Estrato4.Nh)) * Estrato4.Sh2))
Plano4.Vartot[1] = N^2*Plano4.Vartot[1] ## Correção para o estimador de Total de Funcionários
Plano4.Vartot[2] = Vartot.raz
(Plano4.CVtot = 100*sqrt(Plano4.Vartot) / Totais)

rm(aux,Estrato4.Nh,Estrato4.nh,Estrato4.Yh,Estrato4.Sh2,Estrato4.SD )
Resultados[4,] = Plano4.CVtot

##### Plano 5 - Amostragem Aleatória Simples de municípios considerando o emprego do
## estimador de razão tendo como variável auxiliar o total da população.

```

```

X = munic$Populacao
# a) Total de funcionários ativos da administração direta
Y = munic$FuncADMD
Rz = sum(Y)/sum(X)
tot.a = sum(munic$FuncADMD, rm.na=T)

(Plano5.Vartot.a <- N^2 * ((1-f)/n) * sum((Y - Rz*X)^2)*1/(N-1))
(Plano5.CVtot.a = 100*sqrt(Plano5.Vartot.a) / tot.a)

# b) Razão da população por funcionário ativo da administração direta ?????

# c) Proporção de municípios com maternidade
Y = munic$Maternidade
Rz = sum(Y)/sum(X)
prop.c = sum(munic$Maternidade)/N

(Plano5.Vartot.c = ((1-f)/n) * sum((Y - Rz*X)^2)*1/(N-1))
(Plano5.CVtot.c = 100*sqrt(Plano5.Vartot.c)/prop.c)

# d) Proporção de municípios com maternidade e emergência
Y = munic$Mat_Emer
Rz = sum(Y)/sum(X)
prop.d = sum(munic$Mat_Emer)/N

(Plano5.Vartot.d <- ((1-f)/n) * sum((Y - Rz*X)^2)*1/(N-1))
(Plano5.CVtot.d <- 100*sqrt(Plano5.Vartot.d)/prop.d)

Resultados[5,] = c(Plano5.CVtot.a,NA,Plano5.CVtot.c,Plano5.CVtot.d)

#####
#Efeitos Plano Amostral
EPA = Resultados
EPA = EPA[-1,]
for(i in 1:4){
  EPA[i,] <- Resultados[i+1,] / Resultados[1,]
}
#####

round(Resultados,5)
round(EPA,5)

write.csv2(Resultados, file="Resultados.csv")
write.csv2(EPA, file="EPA.csv")

##### SELEÇÃO DA AMOSTRA USANDO PACOTE CONFORME AULA 07 SLIDE 3
(munic.amo = (data=munic, stratanames=c("Estrato3"),Estrato3.nh,method=c("srswor")))
(amostra<-getdata(munic,munic.amo)) #comando para associar dois arquivos
str(amostra)

amostra = svydesign(data=amostra, ids=~1, strata=~Estrato3, fpc=~Prob)
amostra.est_tot = svytotal(~FuncADMD, design=amostra)
amostra.est_raz = svyratio(~Populacao,~FuncADMD, design=amostra)
amostra.est_prop = svymean(~Maternidade+Mat_Emer, design=amostra)

###Parametros
Totais
###Estimativas pontuais
amostra.est_tot
amostra.est_raz
amostra.est_prop

### CV's
(amostra.cv_tot = 100* sqrt(diag(vcov(amostra.est_tot))) / coef(amostra.est_tot))
(amostra.cv_raz = 100* sqrt(diag(vcov(amostra.est_raz))) / coef(amostra.est_raz))
(amostra.cv_prop = 100* sqrt(diag(vcov(amostra.est_prop))) / coef(amostra.est_prop))

```