

**EXAME FINAL**

**Disciplina:** Amostragem

**Professor:** Pedro Luis do Nascimento Silva

**Aluno:** Leandro Lins Marino ([leandromarino@leandromarino.com.br](mailto:leandromarino@leandromarino.com.br))

## **1 Introdução**

O presente trabalho visa comparar os diferentes planos amostrais sob alguns parâmetros populacionais de forma a definir qual o melhor entre eles. A unidade de investigação é o município (5565 municípios) e é fornecida uma base cadastral contendo todas as informações necessárias para as comparações.

Cabe salientar que devido à problemas no cadastro, foram retirados os municípios que se recusaram ou a informação foi ignorada para quaisquer uma das variáveis em estudo. Os municípios retirados foram: Palmas (TO), Campina Grande (PB), Mirandiba (PE), Abaré (BA) e Macaé (RJ). Todos os municípios retirados não apresentam informação para o total de funcionários da administração direta e apenas Abaré (BA) se recusou a prestar informações sobre existência de maternidade e de unidade de emergência no município. A população desses municípios se aproxima de 900 mil habitantes e representa 0,45% da população brasileira. Em termos regionais esta ausência de informações representa, 1,46% da população da região Norte, 0,78% da região Nordeste e 0,26% da região Sudeste.

Sejam as variáveis de pesquisa definidas a seguir

- $Y_1$  – Total de funcionários ativos da administração direta;
- $Y_2$  – Existe Maternidade no município?
- $Y_3$  – Existe Unidade de emergência no município?

A presente pesquisa pretende obter estimativas para os seguintes parâmetros populacionais:

- 1) Total de funcionários ativos da administração direta;
- 2) Razão da população por funcionário ativo da administração direta;
- 3) Proporção de municípios com maternidade;
- 4) Proporção de municípios com maternidade e emergência.

Este trabalho pretende avaliar a influência da determinação de alguns planos amostrais considerando as variáveis de pesquisa e os parâmetros populacionais desejados. A seguir são

apresentados os planos amostrais considerados bem como suas estimativas para os parâmetros, a variância teórica e a variância estimada.

## 2 Planos amostrais

### 2.1 Plano 1

Este plano consiste em selecionar uma amostra aleatória simples dos municípios brasileiros.

#### 2.1.1 Total de funcionários ativos da administração direta

Desta forma, define-se:

$$\hat{Y}_1 = \frac{N}{n} \sum_{i \in S} y_{1i}; \quad Var(\hat{Y}_1) = N^2 \left( \frac{1}{n} - \frac{1}{N} \right) S_{Y_1}^2, \quad \text{onde} \quad S_{Y_1}^2 = \frac{1}{N-1} \sum_{i \in U} (y_{1i} - \bar{Y})^2$$

e sua estimativa

$$\widehat{Var}(\hat{Y}_1) = N^2 \left( \frac{1}{n} - \frac{1}{N} \right) s_{Y_1}^2, \quad \text{onde} \quad s_{Y_1}^2 = \frac{1}{n-1} \sum_{i \in S} (y_{1i} - \bar{y})^2$$

onde  $N$  é o total de municípios na população e  $n$  é o número de municípios na amostra.

#### 2.1.2 Razão da população por funcionário ativo da administração direta;

$$\hat{R} = \frac{\sum_{i \in S} y_{4i}}{\sum_{i \in S} y_{1i}}; \quad Var(\hat{R}) = \left( \frac{1-f}{n * \bar{Y}_4^2} \right) \left( \frac{1}{N-1} \right) \sum_{i \in U} (y_{4i} - R y_{1i})^2, \quad \text{onde} \quad f = \frac{n}{N}.$$

$Y_4$  é o total da população. E,

$$\widehat{Var}(\hat{R}) = \left( \frac{1-f}{n * \bar{Y}_4^2} \right) \left( \frac{1}{N-1} \right) \sum_{i \in S} (y_{4i} - \hat{R} y_{1i})^2$$

#### 2.1.3 Proporção de municípios com maternidade (j=1) e com maternidade e emergência (j=2);

$$\hat{Y}_j = \hat{p}_j = \sum_{i \in S} \frac{y_{ji}}{n}; \quad Var(\hat{Y}_j) = \left( \frac{N-n}{N-1} \right) \frac{P_j(1-P_j)}{n}; \quad \forall j \in \{2,3\}$$

$$\widehat{Var}(\hat{Y}_j) = \left( \frac{N-n}{N-1} \right) \frac{\hat{p}_j(1-\hat{p}_j)}{n}$$

### 2.2 Planos 2 e 3

No plano 2 sorteia-se os municípios através de uma amostragem estratificada simples utilizando a região aonde o município se localiza como variável de estratificação e alocação igual, ou seja, todos os estratos terão 40 municípios selecionados.

Para o plano 3 também é feita uma amostragem estratificada simples de municípios, entretanto, a variável de estratificação é definida de tal forma que a população seja dividida em 5 classes tais que a soma da raiz quadrada da população dos municípios em cada classe seja aproximadamente igual. A alocação, similarmente ao plano 2 será igual em todos os estratos.

Sendo assim, para os planos 2 e 3 os estimadores definidos são os mesmos com a modificação apenas da variável de estratificação.

### 2.2.1 Total de Funcionários da administração direta

$$\hat{Y}_1 = \sum_{i \in S} \sum_{h=1}^H \frac{N_h y_{1hi}}{n_h}; \quad Var(\hat{Y}_1) = N^2 \sum_{h=1}^H W_h^2 \left( \frac{1}{n_h} - \frac{1}{N_h} \right) S_{hy_1}^2, \quad \text{onde}$$

$$S_{hy_1}^2 = \frac{1}{N_h - 1} \sum_{i \in U} (y_{1hi} - \bar{Y}_{1h})^2 \text{ e } h \in \{1, \dots, H\} \text{ é o número de estratos.}$$

### 2.2.2 Razão da população por funcionário ativo da administração direta;

Para este estimador considerando a amostra estratificada simples poderíamos adotar duas abordagens uma considerando o estimador da razão separada, onde são calculadas razões para cada um dos estratos que, posteriormente, compõem uma única estimativa ou o estimador da razão combinada aonde ocorre uma estimativa para o total do denominador e para o total do numerador e a seguir é feita a razão destes totais. Neste estudo iremos apenas considerar o estimador da razão combinada.

$$\hat{R} = \frac{\sum_{h=1}^H \bar{y}_{4h}}{\sum_{h=1}^H \bar{y}_{1h}}; \quad Var(\hat{R}) = \sum_{h=1}^H \frac{N_h^2 (1 - f_h)}{n_h} (S_{y_{4h}}^2 + R^2 S_{y_{1h}}^2 - 2R\rho_h S_{y_{4h}} S_{y_{1h}}), \text{ onde } f_h = \frac{n_h}{N_h}.$$

$$S_{y_{4h}}^2 = \frac{1}{N_h - 1} \sum_{i \in U_h} (y_{4hi} - \bar{Y}_{4h})^2, \quad S_{y_{1h}}^2 = \frac{1}{N_h - 1} \sum_{i \in U_h} (y_{1hi} - \bar{Y}_{1h})^2$$

$\rho_h$  é a correlação dentro de cada estrato  $h, h \in \{1, \dots, H\}$ .

$$\rho_h = \rho_h(Y_1, Y_4) = \frac{\sum_{i \in U_h} (y_{1ih} - \bar{y}_{1h})(y_{4ih} - \bar{y}_{4h})}{(N - 1) S_{y_{1h}} S_{y_{4h}}}$$

$$\widehat{Var}(\hat{R}) = \sum_{h=1}^H \frac{N_h^2 (1 - f_h)}{n_h} (s_{y_{4h}}^2 + \hat{R}^2 s_{y_{1h}}^2 - 2\hat{R}\hat{\rho}_h s_{y_{4h}} s_{y_{1h}}), \text{ onde } f_h = \frac{n_h}{N_h}.$$

$$s_{y_{4h}}^2 = \frac{1}{n_h - 1} \sum_{i \in S_h} (y_{4hi} - \bar{y}_{4h})^2, \quad s_{y_{1h}}^2 = \frac{1}{n_h - 1} \sum_{i \in S_h} (y_{1hi} - \bar{y}_{1h})^2$$

$$\hat{\rho}_h = \hat{\rho}_h(y_1, y_4) = \frac{\sum_{i \in s_h} (y_{1ih} - \bar{y}_{1h})(y_{4ih} - \bar{y}_{4h})}{(n-1)s_{y_1h}s_{y_4h}}$$

### 2.2.3 Proporção de municípios com maternidade (j=1) e com maternidade e emergência (j=2);

Como a proporção é uma média da variável indicadora que define a existência ou não do evento, podemos considerar praticamente os mesmos estimadores apresentados na seção 2.2.1 (Total de funcionários da administração direta) uma vez que a média é o total dividido pelo número de indivíduos.

$$\bar{Y}_j = \sum_{i \in s} \sum_{h=1}^H \frac{y_{jhi}}{n_h}; \quad Var(\hat{Y}_1) = \sum_{h=1}^H W_h^2 \left( \frac{1}{n_h} - \frac{1}{N_h} \right) S_{hY_1}^2, \quad \text{onde}$$

$$S_{hY_1}^2 = \frac{1}{N_h - 1} \sum_{i \in U} (y_{jhi} - \bar{Y}_{jh})^2 \quad e \quad h \in \{1, \dots, H\} \quad \text{é o número de estratos.}$$

$$\widehat{Var}(\hat{Y}_1) = \sum_{h=1}^H W_h^2 \left( \frac{1}{n_h} - \frac{1}{N_h} \right) s_{hY_1}^2, \quad \text{onde}$$

$$s_{hY_1}^2 = \frac{1}{n_h - 1} \sum_{i \in h} (y_{jhi} - \bar{y}_{jh})^2 \quad e \quad h \in \{1, \dots, H\} \quad \text{é o número de estratos.}$$

## 2.3 Plano 4

O plano 4 consiste em uma amostra estratificada em dois estratos sendo um deles formado pelas capitais e municípios com mais de 500 mil habitantes sempre presentes na amostra. O estrato certo conta com 44 municípios brasileiros<sup>1</sup> desta forma, 156 outros municípios serão selecionados via amostra aleatória simples.

As estimativas para razão, total e proporção são as mesmas definidas na seção 2.2 (plano 3). Como há um estrato cujos elementos sempre pertencem à amostra não há estimativa de variância e no caso a variância dos estimadores será a mesma definida pela amostra aleatória simples (considerando uma amostra de 156 entre os 5516 municípios restantes).

## 2.4 Plano 5

Neste plano os municípios serão selecionados via amostragem aleatória simples e será feita o uso do estimador do tipo razão (com variável auxiliar o total populacional dos municípios) para a

---

<sup>1</sup> A capital Palmas (TO) não participa da amostra por não haver informação cadastral para alguma das variáveis da pesquisa.

obtenção das estimativas. Neste plano não está definido variância para a razão entre a população e o número de funcionários ativos da administração direta.

#### 2.4.1 Total de Funcionários da administração direta

$$\hat{Y}_1 = \frac{\bar{y}_1}{\bar{x}} X, \text{ onde } X \text{ é o total populacional}$$

$$Var(\hat{Y}_1) = \frac{N^2(1-f)}{n} \left[ \sum_{i \in U} \frac{(y_{1i} - Rx_i)^2}{N-1} \right] e \widehat{Var}(\hat{Y}_1) = \frac{N^2(1-f)}{n} \left[ \sum_{i \in s} \frac{(y_{1i} - \hat{R}x_i)^2}{n-1} \right]$$

#### 2.4.2 Proporção de municípios com maternidade (j=1) e com maternidade e emergência (j=2);

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

$$Var(\hat{Y}_j) = \frac{(1-f)}{n} \left[ \sum_{i \in U} \frac{(y_{ji} - Rx_i)^2}{N-1} \right] e \widehat{Var}(\hat{Y}_j) = \frac{(1-f)}{n} \left[ \sum_{i \in s} \frac{(y_{ji} - \hat{R}x_i)^2}{n-1} \right]$$

### 3 Metodologia para comparação dos planos amostrais

A comparação entre os planos amostrais dar-se-á através dos coeficientes de variação e dos efeitos do plano amostral (EPA). O coeficiente de variação (CV) é definido por:

$$CV(Y) = \frac{\sqrt{Var(Y)}}{Y}$$

e, para garantir uma melhor comparação entre os planos amostrais, consideraremos o EPA como sendo  $EPA_k(Y) = \frac{CV_k(Y)}{CV_{AAS}(Y)}$ , onde  $k$  é o plano amostral considerado. A Tabela 2 apresenta os coeficientes de variação teóricos para as variáveis consideradas no estudo.

**TABELA 2 – COEFICIENTES DE VARIAÇÃO PARA OS PARÂMETROS DE INTERESSE SEGUNDO O PLANO AMOSTRAL ADOTADO**

Parâmetro de interesse	Coeficientes de Variação				
	Plano 1	Plano 2	Plano 3	Plano 4	Plano 5
Total de funcionários da administração Direta	22.3727	24.8264	9.7403	8.7065	22.0536
Razão entre população e funcionários ativos da administração direta	22.0536	24.1192	8.0646	4.8384	--
Proporção de municípios com maternidade	8.4436	9.5289	8.2343	9.5877	41.2896
Proporção de municípios com maternidade e unidade de emergência	9.5375	10.7467	8.9325	10.8195	41.3184

Em uma breve análise da Tabela 2 pode-se notar que o plano 3 apresenta reduções para todas as variáveis de interesse. A Tabela 3 apresenta o efeito do plano amostral para os diversos planos amostrais. Nota-se claramente que o Plano 3 (amostragem estratificada simples usando a raiz quadrada da população como variável de estratificação) apresentou um melhor desempenho que a amostra aleatória simples em todos os parâmetros de interesse, e, por isto o consideraremos como o melhor dentre os planos testados. Este melhor desempenho se dá em consequência da grande correlação existente entre as variáveis estudadas com a variável auxiliar para a estratificação.

**TABELA 3 – EFEITO DO PLANO AMOSTRAL PARA OS PARÂMETROS DE INTERESSE SEGUNDO O PLANO AMOSTRAL ADOTADO**

Parâmetro de interesse	Efeito do Plano Amostral (referência AAS – Plano 1)			
	Plano 2	Plano 3	Plano 4	Plano 5
Total de funcionários da administração Direta	1.1097	0.4354	0.3891	0.9857
Razão entre população e funcionários ativos da administração direta	1.0937	0.3657	0.2194	--
Proporção de municípios com maternidade	1.1285	0.9752	1.1355	4.8901
Proporção de municípios com maternidade e unidade de emergência	1.1268	0.9366	1.1344	4.3322

Após a comparação dos planos amostrais procedeu-se com a obtenção de uma amostra estratificada simples de tamanho n=200. A Tabela 3 apresenta as estimativas, erro padrão e coeficiente de variação para os parâmetros de interesse. Os coeficientes de variação estimados são próximos aos coeficientes de variação teóricos para o plano amostral adotado.

**TABELA 3 – ESTIMATIVAS, ERRO PADRÃO E COEFICIENTE DE VARIAÇÃO PARA AS VARIÁVEIS DE INTERESSE NA AMOSTRA SELECIONADA**

Parâmetros de Interesse	Parâmetro Populacional	Estatísticas		
		Estimativa Pontual	Erro Padrão	Coeficiente de Variação
Total de funcionários da administração Direta	5,637,624	6,610,939	976,886	14.7768
Razão entre população e funcionários ativos da administração direta	33.9704	48.4500	4.7229	9.7481
Proporção de municípios com maternidade	0.4038	0.3190	0.0311	9.7388
Proporção de municípios com maternidade e unidade de emergência	0.3492	0.2925	0.0293	10.0082

Por fim, nota-se que, considerando um intervalo de confiança de 95% para as estimativas pontuais, apenas a razão entre a população e funcionários ativos da administração direta não contém o valor real da população.

## 4 Considerações finais

Considerando a grande correlação ( $\rho = 0.9288$ ) existente entre as o total de funcionários da administração direta e a população dos municípios, talvez um plano amostral que incorporasse probabilidades diferentes de seleção em função do tamanho populacional dos mesmos fornecesse uma estimativa com menor coeficiente de variação.

## 5 Referências Bibliográficas

COCHRAN, W.G. **Sampling Techniques**, 3ª ed. Nova Iorque, EUA: John Wiley & Sons, 1977.

## Anexo A – Código Reprodutível em R

O arquivo contendo este código encontra-se disponível em:

<https://raw.githubusercontent.com/leandromarino/TBAmoMunic/master/TbAmo.r>

```
#####
#####
##### IDENTIFICACAO #####
#####
#####
#####
#####

### DEFINICAO DO TRABALHO E DE VARIAVEIS DE AMBIENTE
options(width=180,scipen=50,repos = 'http://cran-r.c3sl.ufpr.br/')

#instalar pacote a partir de zip local
#dirpacote<-'C:\\Users\\leandromarino\\Documents\\Projetos\\TBAmoMunic\\trunk\\'
#install.packages(paste(dirpacote,'RCurl_1.95-3.zip'))

### CARREGANDO OS PACOTES NECESSARIOS
library(xlsReadWrite)
library(SOAR)
library(survey)
library(stratification)
library(sampling)
require(RCurl)
require(foreign)

### LEITURA DOS DADOS
## Inicialmente o arquivo estava em formato .xlsx e foi convertido no Excel 2010
## para o .xls
url <- 'https://raw.githubusercontent.com/leandromarino/TBAmoMunic/master/Munic.dat'
url <- getURL(url,ssl.verifypeer = FALSE)
munic <- read.table(textConnection(url), colClasses='character',header=T,quote='',
  sep='\t')
str(munic)

pla.cv <- plb.cv <- plc.cv <- pld.cv <-
p2a.cv <- p2b.cv <- p2c.cv <- p2d.cv <-
p3a.cv <- p3b.cv <- p3c.cv <- p3d.cv <-
p4a.cv <- p4b.cv <- p4c.cv <- p4d.cv <-
p5a.cv <- p5b.cv <- p5c.cv <- p5d.cv <- NA

#Y1 = FuncADMD - Total de funcionario ativos da administracao direta;
#Y2 = Maternidade - Existe Maternidade no municipio?
#Y3 = Emergencia - Existe Unidade de emergencia no municipio?
munic <- transform(munic,
  Populacao=as.integer(Populacao),
  FuncADMD=as.integer(FuncADMD),
  Regiao=substr(CodMunic,1,1))
munic <- transform(munic,
  mater = ifelse(Maternidade=="Sim",1,ifelse(Maternidade=="Não",0,NA)),
  emerg = ifelse(Emergencia=="Sim",1,ifelse(Emergencia=="Não",0,NA)))
munic <- transform(munic,
  materemerg = ifelse(rowSums(cbind(mater,emerg)) == 2,1,0))
munic.exc <- munic[rowSums(is.na(munic))!=0,]
munic <- munic[rowSums(is.na(munic))==0,]
dim(munic)
munic[1:10,]

for( i in 1:5){
```



```

print( sum(munic.exc[munic.exc$Regiao==i, 'Populacao']) /
      (sum(munic.exc[munic.exc$Regiao==i, 'Populacao']) +
        sum(munic[munic$Regiao==i, 'Populacao'])) *100)
}

#####-----#####
#####-----#####
### ITEM 1 - Use os valores populacionais das variaveis de interesse para
### determinar o CV esperado para os estimadores dos parametros de interesse.

### PLANO 1 - AAS
### PLANO 2 - AES de Munic por Reg com Alloc igual
### PLANO 3 - AES de Munic por Pop talque sqrt(Estrado_i) == sqrt(Estrato_j)
### \forall i,j \in 1:5
### PLANO 4 - AES por 'corte' pi = 1 se pop >= 500.000 e cap, aas nos demais
### PLANO 5 - AAS com uso do estimador de razao usando var aux o tot da pop

# Considere de interesse estimar os seguintes parâmetros populacionais:
# a) Total de funcionários ativos da administração direta;
# b) Razão da população por funcionário ativo da administração direta;
# c) Proporção de municípios com maternidade;
# d) Proporção de municípios com maternidade e emergência.
# Considere a idéia de selecionar uma amostra de n=200 municípios para uma
# pesquisa por amostragem junto à população de municípios, com um dos planos
# amostrais abaixo indicados.

N <- nrow(munic)
n <- 200
(f <- n/N)

#####
## Plano 1
#####
# a) Total de funcionários ativos da administração direta
(pla.var <- N^2 * ( 1/n - 1/N ) * var(munic$FuncADMD))
(pla.tot <- sum(munic$FuncADMD))
(pla.cv <- sqrt(pla.var)/pla.tot)

# b) Razão da população por funcionário ativo da administração direta;
med = mean(munic$FuncADMD)
(plb.raz <- sum(munic$Populacao)/sum(munic$FuncADMD))
(plb.var <- (1-f) / (n*med^2)*1/(N-1)*sum((munic$Populacao-plb.raz*munic$FuncADMD)^2))
(plb.cv <- sqrt(plb.var)/plb.raz)

# c) Proporção de municípios com maternidade;
(plc.prop <- mean(munic$mater))
(plc.var <- ((N-n)/(N-1)) * plc.prop * (1-plc.prop)*(1/n))
(plc.cv <- sqrt(plc.var)/plc.prop)

# d) Proporção de municípios com maternidade e emergência.
(pld.prop <- mean(munic$materemerg))
(pld.var <- ((N-n)/(N-1)) * pld.prop * (1-pld.prop)*(1/n))
(pld.cv <- sqrt(pld.var)/pld.prop)

#####
## Plano 2
#####
estrato <- munic$Regiao
(Nest <- matrix(table(estrato),ncol=5))
(nest <- rep(200/5,5))
(fest <- nest/Nest)
(Wh <- Nest/N)

est.munic <- split(munic,factor(estrato))

# a) Total de funcionários ativos da administração direta

```

```

func <- split(munic$FuncADMD,estrato)
(tot.est <- do.call(c,lapply(func,sum)))
(p2a.tot <- sum(tot.est))
(var.intra <- do.call(c,lapply(func,var)))
(p2a.var <- N^2 * sum(Wh^2 * (1/nest - 1/Nest)*var.intra))
(p2a.cv <- sqrt(p2a.var)/p2a.tot)

# b) Razão da população por funcionário ativo da administração direta;
func <- split(munic$FuncADMD,estrato)
pop <- split(munic$Populacao,estrato)
(p2b.raz <- sum(munic$Populacao)/sum(munic$FuncADMD))
(raz.est <- p2b.raz)
(medest <- do.call(c,lapply(func,mean)))
(X <- sum(munic$FuncADMD))
(var.intraX <- do.call(c,lapply(func,var)))
(var.intraY <- do.call(c,lapply(pop,var)))
(sd.intraX <- sqrt(var.intraX))
(sd.intraY <- sqrt(var.intraY))
cor.XY <- list()
for(i in 1:5) cor.XY[[i]] <- cor(func[[i]],pop[[i]])
(cor.XY <- do.call(c,cor.XY))
(p2b.var <- (1/X^2) * sum(Nest^2 * (1- fest)/nest * (var.intraY +
raz.est^2*var.intraX - 2 * raz.est * sd.intraY * sd.intraX * cor.XY)))
(p2b.cv <- sqrt(p2b.var)/p2b.raz)

# c) Proporção de municípios com maternidade;
mater <- split(munic$mater,estrato)
(prop.est <- do.call(c,lapply(mater,mean)))
(p2c.prop <- sum(Wh*prop.est))
(var.intra <- do.call(c,lapply(mater,var)))
(p2c.var <- sum(Wh^2 * (1/nest - 1/Nest)*var.intra))
(p2c.cv <- sqrt(p2c.var)/p2c.prop)

# d) Proporção de municípios com maternidade e emergência.
materemerg <- split(munic$materemerg,estrato)
(prop.est <- do.call(c,lapply(materemerg,mean)))
(p2d.prop <- sum(Wh*prop.est))
(var.intra <- do.call(c,lapply(materemerg,var)))
(p2d.var <- sum(Wh^2 * (1/nest - 1/Nest)* var.intra))
(p2d.cv <- sqrt(p2d.var)/p2d.prop)

#####
## Plano 3
#####
sum(munic$Populacao)/5
munic <- transform(munic,sqrtPop = sqrt(Populacao))
munic <- munic[order(munic$sqrtPop),]

sum(munic$sqrtPop)/5

aux <- matrix(NA,nrow=nrow(munic),ncol=5)
for(i in 1:5){
aux[,i] <- cumsum(munic$sqrtPop) <= (sum(munic$sqrtPop)/5)*i
}
estrato <- 6-rowSums(aux)
table(estrato)
for(i in 1:5) print(sum(munic$sqrtPop[estrato==i]))

(Nest <- matrix(table(estrato),ncol=5))
(nest <- rep(200/5,5))
(Wh <- Nest/N)
est.munic <- split(munic,factor(estrato))

# a) Total de funcionários ativos da administração direta
func <- split(munic$FuncADMD,estrato)

```

```

(tot.est <- do.call(c,lapply(func,sum)))
(p3a.tot <- sum(tot.est))
(var.intra <- do.call(c,lapply(func,var)))
(p3a.var <- N^2 * sum(Wh^2 * (1/nest - 1/Nest)*var.intra))
(p3a.cv <- sqrt(p3a.var)/p3a.tot)

# b) Razão da população por funcionário ativo da administração direta;
func <- split(munic$FuncADMD,estrato)
pop <- split(munic$Populacao,estrato)
(p3b.raz <- sum(munic$Populacao)/sum(munic$FuncADMD))
(raz.est <- p3b.raz)
(medest <- do.call(c,lapply(func,mean)))
(X <- sum(munic$FuncADMD))
(var.intraX <- do.call(c,lapply(func,var)))
(var.intraY <- do.call(c,lapply(pop,var)))
(sd.intraX <- sqrt(var.intraX))
(sd.intraY <- sqrt(var.intraY))
cor.XY <- list()
for(i in 1:5) cor.XY[[i]] <- cor(func[[i]],pop[[i]])
(cor.XY <- do.call(c,cor.XY))
(p3b.var <- (1/X^2) * sum(Nest^2 * (1- fest)/nest * (var.intraY +
raz.est^2*var.intraX - 2 * raz.est * sd.intraY * sd.intraX * cor.XY)))
(p3b.cv <- sqrt(p3b.var)/p3b.raz)

# c) Proporção de municípios com maternidade;
mater <- split(munic$mater,estrato)
(prop.est <- do.call(c,lapply(mater,mean)))
(p3c.prop <- sum(Wh*prop.est))
(var.intra <- do.call(c,lapply(mater,var)))
(p3c.var <- sum(Wh^2 * (1/nest - 1/Nest)*var.intra))
(p3c.cv <- sqrt(p3c.var)/p3c.prop)

# d) Proporção de municípios com maternidade e emergência.
materemerg <- split(munic$materemerg,estrato)
(prop.est <- do.call(c,lapply(materemerg,mean)))
(p3d.prop <- sum(Wh*prop.est))
(var.intra <- do.call(c,lapply(materemerg,var)))
(p3d.var <- sum(Wh^2 * (1/nest - 1/Nest)* var.intra))
(p3d.cv <- sqrt(p3d.var)/p3d.prop)

#####
## Plano 4
#####
munic <- munic[order(munic$CodMunic),]
capitais <- c('1100205','1200401','1302603','1400100','1501402','1600303',
'1721000','2111300','2211001','2304400','2408102','2507507',
'2611606','2704302','2800308','2927408','3106200','3205309',
'3304557','3550308','4106902','4205407','4314902','5002704',
'5103403','5208707','5300108')
estratocerto <- rep(FALSE,nrow(munic))
estratocerto[is.element(munic$CodMunic,substr(capitais,1,6))] <- TRUE
estratocerto[munic$Populacao >= 500000] <- TRUE
table(estratocerto)

estrato <- rep(0,nrow(munic))
estrato[ estratocerto] <- 1
estrato[!estratocerto] <- 2

aux <- rep(FALSE,nrow(munic))
aux[estrato==2] <- TRUE

(Nest <- table(estrato)[2])
(nest <- 200-44)
est.munic <- split(munic,factor(estrato))

```

```

N <- Nest
n <- nest
(f <- n/N)

# a) Total de funcionários ativos da administração direta
(p4a.var <- N^2 * ( 1/n - 1/N ) * var(munic$FuncADMD[aux]))
(p4a.tot <- sum(munic$FuncADMD))
(p4a.cv <- sqrt(p4a.var)/p4a.tot)

# b) Razão da população por funcionário ativo da administração direta;
med = mean(munic$FuncADMD)
(p4b.raz <- sum(munic$Populacao)/sum(munic$FuncADMD))
(razao.est <- sum(munic$Populacao[aux])/sum(munic$FuncADMD[aux]))
(p4b.var <- (1-f)/(n*med^2)*1/(N-1)*
  sum((munic$Populacao[aux]-razao.est*munic$FuncADMD[aux])^2))
(p4b.cv <- sqrt(p4b.var)/p4b.raz)

# c) Proporção de municípios com maternidade;
(p4c.prop <- mean(munic$mater))
(prop.est <- mean(munic$mater[aux]))
(p4c.var <- ((N-n)/(N-1)) * prop.est * (1-prop.est)*(1/n))
(p4c.cv <- sqrt(p4c.var)/p4c.prop)

# d) Proporção de municípios com maternidade e emergência.
(p4d.prop <- mean(munic$materemerg))
(prop.est <- mean(munic$materemerg[aux]))
(p4d.var <- ((N-n)/(N-1)) * prop.est * (1-prop.est)*(1/n))
(p4d.cv <- sqrt(p4d.var)/p4d.prop)

#####
## Plano 5
#####

N <- nrow(munic)
n <- 200
(f <- n/N)

(medX <- mean(munic$Populacao))
X <- munic$Populacao
Y <- munic$FuncADMD
(R <- sum(Y)/sum(X))

# a) Total de funcionários ativos da administração direta
(p5a.var <- N^2 * ((1-f)/n) * sum((Y - R*X)^2)*1/(N-1))
(p5a.tot <- sum(munic$FuncADMD))
(p5a.cv <- sqrt(p5a.var)/p5a.tot)

# b) Razão da população por funcionário ativo da administração direta;
#Este cálculo não faz sentido.

# c) Proporção de municípios com maternidade;
Y <- munic$mater
(R <- sum(Y)/sum(X))
(p5c.var <- ((1-f)/n) * sum((Y - R*X)^2)*1/(N-1))
(p5c.tot <- mean(munic$mater))
(p5c.cv <- sqrt(p5c.var)/p5c.tot)

# d) Proporção de municípios com maternidade e emergência.
Y <- munic$materemerg
(R <- sum(Y)/sum(X))
(p5d.var <- ((1-f)/n) * sum((Y - R*X)^2)*1/(N-1))
(p5d.tot <- mean(munic$materemerg))
(p5d.cv <- sqrt(p5d.var)/p5d.tot)

```

```

resumo <- data.frame(plano = paste('Plano',1:5),
                    parametro_a=c(pla.cv,p2a.cv,p3a.cv,p4a.cv,p5a.cv)*100,
                    parametro_b=c(plb.cv,p2b.cv,p3b.cv,p4b.cv,p5b.cv)*100,
                    parametro_c=c(plc.cv,p2c.cv,p3c.cv,p4c.cv,p5c.cv)*100,
                    parametro_d=c(pld.cv,p2d.cv,p3d.cv,p4d.cv,p5d.cv)*100)

resumo

epas <- resumo[-1,]
for(i in 1:4){
  epas[i,2:5] <- resumo[i+1,2:5] / resumo[1,2:5]
}
epas

#write.xls(resumo,'c:/Projetos/TbAmoMunic/trunk/cvs.xls')
#write.xls(epas,'c:/Projetos/TbAmoMunic/trunk/epas.xls')

#####-----#####
#####-----#####
### ITEM 3 - Selecione uma amostra de municípios segundo o esquema amostral que
###você escolheu em 2.

sum(munic$Populacao)/5
munic <- transform(munic,sqrtPop = sqrt(Populacao))
munic <- munic[order(munic$sqrtPop),]
sum(munic$sqrtPop)/5
aux <- matrix(NA,nrow=nrow(munic),ncol=5)
for(i in 1:5){
  aux[,i] <- cumsum(munic$sqrtPop) <= (sum(munic$sqrtPop)/5)*i
}
estrato <- 6-rowSums(aux)
table(estrato)

munic$estrato <- estrato
munic.est <- split(munic,estrato)

munic.amo <- list()
for( i in 1:5){
  aux <- sample(1:nrow(munic.est[[i]]),40)
  munic.amo[[i]] <- munic.est[[i]][aux,]
}

amo.munic <- do.call(rbind,munic.amo)
dim(amo.munic)
str(amo.munic)

#write.table(amo.munic,'c:/Projetos/TbAmoMunic/trunk/amo_Munic.dat',sep='\t',
#            quote=F,row.names=F)
url <- 'https://raw.githubusercontent.com/leandromarino/TBAmoMunic/master/amo_Munic.dat'
url <- getURL(url,ssl.verifypeer = FALSE)
amo.munic <- read.table(textConnection(url), colClasses='character',header=T,quote='',
                        sep='\t')
str(amo.munic)

amo.munic <- transform(amo.munic,
                      Populacao=as.integer(Populacao),
                      FuncADMD=as.integer(FuncADMD),
                      Regiao=as.integer(Regiao),
                      mater = as.integer(mater),
                      emerg = as.integer(emerg),
                      materemerg = as.integer(materemerg),
                      sqrtPop = as.numeric(sqrtPop))
amo.munic[1:10,]
str(amo.munic)

```

```

#####-----#####
#####-----#####
### ITEM 4 - De posse da amostra de municípios selecionada, estime os parâmetros
###de interesse e seus respectivos erros padrão e coeficientes de variação

#####

(Nest <- matrix(table(estrato),ncol=5))
(N = sum(Nest))
(nest <- rep(200/5,5))
(Wh <- Nest/N)
est.munic <- munic.amo
estrato <- amo.munic$estrato

# a) Total de funcionários ativos da administração direta
func <- split(amo.munic$FuncADMD,estrato)
(tot.est <- do.call(c,lapply(func,sum)))
(amop3a.tot <- sum(tot.est/nest*Nest))
(var.intra <- do.call(c,lapply(func,var)))
(amop3a.var <- N^2 * sum(Wh^2 * (1/nest - 1/Nest)*var.intra))
(amop3a.cv <- sqrt(amop3a.var)/amop3a.tot)

# b) Razão da população por funcionário ativo da administração direta;
func <- split(amo.munic$FuncADMD,estrato)
pop <- split(amo.munic$Populacao,estrato)
#(amop3b.raz <- sum(amo.munic$Populacao)/sum(amo.munic$FuncADMD))
(amop3b.raz <- sum(do.call(c,lapply(pop,sum))/40) /
sum(do.call(c,lapply(func,sum))/40))
(raz.est <- amop3b.raz)
(medest <- do.call(c,lapply(func,mean)))
(X <- sum(medest*Nest))
(var.intraX <- do.call(c,lapply(func,var)))
(var.intraY <- do.call(c,lapply(pop,var)))
(sd.intraX <- sqrt(var.intraX))
(sd.intraY <- sqrt(var.intraY))
cor.XY <- list()
for(i in 1:5) cor.XY[[i]] <- cor(func[[i]],pop[[i]])
(cor.XY <- do.call(c,cor.XY))
(amop3b.var <- (1/X^2) * sum(Nest^2 * (1- fest)/nest * (var.intraY +
raz.est^2*var.intraX - 2 * raz.est * sd.intraY * sd.intraX * cor.XY)))
(amop3b.cv <- sqrt(amop3b.var)/amop3b.raz)

# c) Proporção de amo.municípios com maternidade;
mater <- split(amo.munic$mater,estrato)
(prop.est <- do.call(c,lapply(mater,mean)))
(amop3c.prop <- sum(Wh*prop.est))
(var.intra <- do.call(c,lapply(mater,var)))
(amop3c.var <- sum(Wh^2 * (1/nest - 1/Nest)*var.intra))
(amop3c.cv <- sqrt(amop3c.var)/amop3c.prop)

# d) Proporção de amo.municípios com maternidade e emergência.
materemerg <- split(amo.munic$materemerg,estrato)
(prop.est <- do.call(c,lapply(materemerg,mean)))
(amop3d.prop <- sum(Wh*prop.est))
(var.intra <- do.call(c,lapply(materemerg,var)))
(amop3d.var <- sum(Wh^2 * (1/nest - 1/Nest)* var.intra))
(amop3d.cv <- sqrt(amop3d.var)/amop3d.prop)

resumo.amo <- rbind(
c(amop3a.tot ,sqrt(amop3a.var),amop3a.cv*100),
c(amop3b.raz ,sqrt(amop3b.var),amop3b.cv*100),
c(amop3c.prop,sqrt(amop3c.var),amop3c.cv*100),
c(amop3d.prop,sqrt(amop3d.var),amop3d.cv*100))

#write.xls(resumo.amo,'c:/Projetos/TbAmoMunic/trunk/resumo_amostra.xls')

```