



Diretoria de Pesquisas - DPE
Coordenação de Trabalho e Rendimento - COREN

**ESTIMAÇÃO DE INTERVALOS DE CONFIANÇA PARA ESTIMADORES DE DIFERENÇAS
TEMPORAIS NA PESQUISA MENSAL DE EMPREGO.**

(Versão preliminar)

Maurício Franca Lila

Marcos Paulo Soares de Freitas

RIO DE JANEIRO
17 de Agosto de 2003

1. Introdução

O objetivo deste trabalho foi fornecer informações estatísticas a respeito de um conjunto de indicadores da Pesquisa Mensal de Emprego que pudessem sustentar indicações de mudanças na conjuntura econômica.

A metodologia utilizada consiste em calcular intervalos de confiança para a diferença temporal para um determinado conjunto de indicadores provenientes da pesquisa, para cada região metropolitana isoladamente e para o conjunto das seis.

A questão de diferenças entre estimadores de taxas para uma determinada Região Metropolitana foi tratada por Corrêa, Freitas e Silva (2001), mostrando que conclusões equivocadas haviam sido feitas devido ao fato de se ignorar o processo de amostragem da pesquisa no momento de realizar análises acerca dos resultados da taxa de desocupação.

Neste estudo foram calculados intervalos de confiança para diferenças entre estimativas de 41 indicadores, considerando intervalos de tempo mensais, trimestrais e semestrais da pesquisa.

Primeiramente será descrito o processo de estimação de diferenças entre totais para uma dada região metropolitana. Posteriormente será abordado o tema de estimação de diferenças entre taxas para o conjunto das seis Regiões Metropolitanas da pesquisa.

2. A Pesquisa Mensal de Emprego

A Pesquisa Mensal de Emprego (PME) é uma pesquisa por amostra de unidades domiciliares que produz indicadores referentes ao mercado de trabalho com periodicidade mensal para a área urbana de um conjunto de seis regiões

metropolitanas(Recife, Salvador, Belo Horizonte, Rio de Janeiro, São Paulo e Porto Alegre) e para cada região isoladamente. A população alvo da pesquisa é constituída por pessoas com 10 anos ou mais de idade.

A PME adota um plano amostral estratificado e conglomerado em um ou dois estágios de seleção dependendo do estrato de seleção.

A estratificação na PME ocorre em duas etapas geográficas, a primeira estratifica as unidades primárias de amostragem por Região Metropolitana da pesquisa, A segunda etapa de estratificação ocorre dentro de cada Região Metropolitana é feita a estratificação das unidades de amostragem por municípios ou pseudo-municípios¹.

As unidades primárias de amostragem são os setores censitários que correspondem a unidade de controle cadastral formada por área contínua, situada em um único quadro urbano ou rural. No caso desta pesquisa são selecionados somente os setores urbanos. A seleção das unidades primárias de amostragem é feita através de amostragem sistemática proporcional ao total de domicílios ocupados encontrados no setor censitário, na ocasião do Censo Demográfico 2000.

Após a seleção dos setores censitários é feita então por meio de amostragem sistemática simples, a seleção das unidades secundárias de amostragem que são os domicílios.

A metodologia de coleta utilizada na PME pesquisa um domicílio por 4 meses consecutivos, após este período o domicílio permanece oito meses sem ser investigado, e volta a ser pesquisado após este período por mais quatro meses, sendo excluído definitivamente da amostra da pesquisa.

A amostra mensal da PME é subdividida em 8 grupos de rotação, onde 25% da amostra de domicílios é substituída a cada mês, seguindo um esquema de grupos de

¹ Correspondem a agregados de municípios de menor porte criados de maneira que um número mínimo de unidades de amostragem sejam selecionadas dentro de cada estrato de seleção da pesquisa.

rotação e painéis, onde cada painel corresponde a um conjunto de domicílios e cada grupo de rotação corresponde a um conjunto de setores censitários selecionados.

Com o objetivo de controlar os crescimentos significativos de unidades domiciliares em certos setores censitários da amostra, o que pode ocorrer ao longo da década, é realizado o cadastro de "novas construções", que identifica esses crescimentos para que se possa dar um tratamento diferenciado a estas partes.

O universo de novas construções é estratificado por município, sendo cada domicílio selecionado segundo a fração amostral adotada para a área correspondente.

O quadro 1 mostra as frações amostrais adotadas para cada área de abrangência da pesquisa.

Quadro 1 – Fração amostral da PME por área de abrangência da pesquisa

Área de abrangência	Fração amostral
Recife	1/200
Salvador	1/200
Belo Horizonte	1/200
Rio de Janeiro	1/500
São Paulo	1/700
Porto Alegre	1/200

Fonte: Pesquisa Mensal de Emprego, Série Relatórios Metodológicos, IBGE.

O fator de ponderação da pesquisa é corrigido pelo tratamento de não resposta.

Para evitar que domicílios localizados em áreas onde a taxa de não entrevista havia sido alta recebessem um peso muito alto, foi estabelecido que o fator de ponderação não poderia ultrapassar 5 vezes o valor da fração amostral adotada para a área correspondente. A pesquisa também utiliza a projeção de população para atualizar suas estimativas.

As estimativas de variância para os resultados da PME são obtidos através do método do *Ultimate Cluster*, também conhecido como método do Conglomerado Primário (Cochran, p.307) utilizando estimadores de razão.

3. A Linearização de Taylor no processo de estimação de variância.

Como as quantidades de interesse na PME são obtidas através de estimadores de razão obtido por valores estimados através do processo de expansão da amostra, não existe uma fórmula exata para as variâncias destas quantidades. No entanto pode-se obter uma aproximação da variância de uma razão através de Linearização de Taylor. Este método consiste em definir uma variável linearizada z e aproximar a variância da estatística amostral de interesse pela variância dos valores z sob o plano amostral utilizado para a seleção da amostra.

$$\text{Sejam } F(X, Y) = \frac{Y}{X} = R \quad (3.1) \text{ e } F(\hat{X}, \hat{Y}) = \frac{\hat{Y}}{\hat{X}} = \hat{R} \quad (3.2).$$

Onde R representa a razão entre duas quantidades de interesse da pesquisa. O estimador \hat{R} pode ser escrito da seguinte forma:

$$\hat{R} = F(X, Y) \partial F_x(X, Y)(\hat{X} - X) + F(X, Y) \partial F_y(X, Y)(\hat{Y} - Y) + \text{termos de ordem superior} \quad (3.3)$$

$$\hat{R} \cong F(X, Y) \partial F_x(X, Y)(\hat{X} - X) + F(X, Y) \partial F_y(X, Y)(\hat{Y} - Y) \quad (3.4)$$

O erro quadrático médio de \hat{R} é determinado através de :

$$EQM(\hat{R}) = [\hat{R} - F(X, Y)]^2 \quad (3.5)$$

$$[\hat{R} - F(X, Y)]^2 \cong (\partial F_x(X, Y))^2 V(\hat{X}) + (\partial F_y(X, Y))^2 V(\hat{Y}) + 2(\partial F_x(X, Y))(\partial F_y(X, Y)) \text{COV}(\hat{X}, \hat{Y}) \quad (3.6)$$

Fazendo $\hat{Z} = (\partial F_x(X, Y))\hat{X} + (\partial F_y(X, Y))\hat{Y}$ (3.7), tem-se:

$$EQM(\hat{R}) \equiv V[(\partial F_x(X, Y))\hat{X} + (\partial F_y(X, Y))\hat{Y}] = V(\hat{Z}) \quad (3.8)$$

Portanto para estimar a variância da razão \hat{R} torna-se necessário a construção da variável \hat{Z} como foi descrito anteriormente.

4. Estimador da variância do estimador da diferença de totais obtidos considerando pós-estratificação para cada uma das seis Regiões metropolitanas.

O estimador natural do desenho para o total de uma determinada característica de interesse na pesquisa pode ser descrito através do seguinte estimador:

$$\hat{Y}_g^t = \sum_{h=1}^{H_g} \sum_{i=1}^{n_{gh}} \sum_{j=1}^{m_{ghi}} \sum_{k=1}^{m_{ghj}} \frac{1}{f_g} y_{ghijk} = \sum_{h=1}^{H_g} \frac{1}{n_{gh}} \sum_{i=1}^{n_{gh}} \frac{1}{p_{ghi}} \frac{M_{ghi}}{m_{ghi}} \sum_{j=1}^{m_{ghi}} \sum_{k=1}^{m_{ghj}} \frac{1}{f_g} y_{ghijk} \quad (4.1)$$

Onde:

f_g corresponde a fração de amostragem utilizada numa determinada Região Metropolitana g ;

H_g corresponde ao número de estratos de seleção na Região Metropolitana g ;

n_{gh} é o número de setores selecionados no h -ésimo estrato de seleção da Região Metropolitana g ;

m_{ghi} é o número de domicílios selecionados na amostra no i -ésimo setor censitário do h -ésimo estrato de seleção da Região Metropolitana g ;

m_{ghij} é o número de indivíduos selecionados na amostra, no j -ésimo domicílio no i -ésimo setor censitário do h -ésimo estrato de seleção da Região Metropolitana g ;

M_{ghi} é o número de domicílios (ocupados + fechados + vagos + ocasionais + unidades de habitação em domicílios coletivos) contabilizados através da listagem atualizada no i -ésimo setor censitário do h -ésimo estrato de seleção da Região Metropolitana g ;

$p_{ghi} = \frac{M_{ghi}^o}{M_{gh}^o}$ que corresponde ao tamanho relativo do i -ésimo setor, no h -ésimo estrato da Região Metropolitana g , na ocasião do Censo Demográfico de 2000.

y_{ghijk} corresponde ao valor da variável de interesse y para k -ésima pessoa no j -domicílio no i -ésimo setor censitário do h -ésimo estrato de seleção da Região Metropolitana g ;

Como a pesquisa utiliza o estimador combinado de razão, o estimador do total com calibração pela população projetada para o instante de tempo é descrito da seguinte maneira:

$$\hat{Y}_{rg}^t = \hat{Y}_g^t \frac{P_g^t}{\hat{P}_g^t} \quad (4.2)$$

Portanto, a diferença entre dois totais para dois instantes distintos de tempo, pode ser obtida a partir da seguinte equação:

$$\hat{Y}_{rg}^t - \hat{Y}_{rg}^{t-k} = \hat{Y}_g^t \frac{P_g^t}{\hat{P}_g^t} - \hat{Y}_g^{t-k} \frac{P_g^{t-k}}{\hat{P}_g^{t-k}} \quad (4.3)$$

Onde o respectivo estimador de variância desta quantidade é dado por:

$$\hat{V}(\hat{Y}_{rg}^t - \hat{Y}_{rg}^{t-k}) = \hat{V}(\hat{Y}_{rg}^t) + \hat{V}(\hat{Y}_{rg}^{t-k}) - 2COV(\hat{Y}_{rg}^t, \hat{Y}_{rg}^{t-k}) \quad (4.4)$$

Neste caso o estimador de variância $\hat{V}(\hat{Y}_{rg}^t)$ é obtido através de uma aproximação pelo Método de Linearização de Taylor, sendo o mesmo processo válido para o cálculo da parcela de covariância.

Por analogia ao processo descrito na seção 3 temos:

$$\hat{Y}_{rg}^t = \hat{Y}_g^t \frac{P_g^t}{\hat{P}_g^t} = F(\hat{Y}_g^t, \hat{P}_g^t) \quad (4.5)$$

Onde as derivadas parciais em relação a cada variável correspondem a:

$$\frac{\partial F(\hat{Y}_g^t, \hat{P}_g^t)}{\partial \hat{Y}_g^t} = \frac{P_g^t}{\hat{P}_g^t} \quad (4.6)$$

$$\frac{\partial F(\hat{Y}_g^t, \hat{P}_g^t)}{\partial \hat{P}_g^t} = -\hat{Y}_g^t \frac{P_g^t}{(\hat{P}_g^t)^2} \quad (4.7)$$

Portanto neste caso a variável linearizada Z_{ghijk}^t assume o seguinte valor:

$$z_{ghijk}^t = \left(\frac{P_g^t}{\hat{P}_g^t} y_{ghijk} - \hat{Y}_g^t \frac{P_g^t}{(\hat{P}_g^t)^2} P_{ghijk} \right) w_{ghijk} \quad (4.8)$$

Onde w_{ghijk} corresponde ao fator de expansão da k-ésima pessoa no j-domicílio no h-ésimo setor censitário do h-ésimo estrato de seleção da Região Metropolitana g;

Assumindo, $\hat{R}_g^t = \frac{\hat{Y}_g^t}{\hat{P}_g^t}$ temos a seguinte expressão alternativa para (4.8).

$$z_{ghijk}^t = P_g^t \left(\frac{y_{ghijk} - \hat{R}_g^t P_{ghijk}}{\hat{P}_g^t} \right) w_{ghijk} .$$

(4.9)

Portanto a variância da diferença entre dois totais obtidos em cada Região Metropolitana da pesquisa entre instantes de tempo distintos pode ser representada por:

$$\begin{aligned} \hat{V}(\hat{Y}_{rg}^t - \hat{Y}_{rg}^{t-k}) \cong & \sum_{h=1}^{H_g} \sum_{i=1}^{n_{gh}^t} \frac{(z_{ghi}^t - \bar{z}_{gh}^t)^2}{n_{gh}^t - 1} + \\ & + \sum_{h=1}^{H_g} \sum_{i=1}^{n_{gh}^{t-k}} \frac{(z_{ghi}^{t-k} - \bar{z}_{gh}^{t-k})^2}{n_{gh}^{t-k} - 1} - 2 \sum_{h=1}^{H_g} \sum_{i=1}^{n_{gh}^*} \frac{(z_{ghi}^{t^*} - \bar{z}_{gh}^{t^*})(z_{ghi}^{(t-k)^*} - \bar{z}_{gh}^{(t-k)^*})}{n_{gh}^* - 1} \end{aligned} \quad (4.10)$$

Onde:

$$z_{ghi}^t = \sum_{j=1}^{m_{ghi}} \sum_{k=1}^{m_{ghj}} z_{ghijk}^t \quad (4.11)$$

$$\bar{z}_{gh}^t = \frac{1}{n_{gh}^t} \sum_{j=1}^{n_{hi}} z_{ghi}^t \quad (4.12)$$

$z_{ghi}^{t^*}$ representa o valor da variável z no instante t que possui um par ordenado no instante $t - k$ e n_{gh}^* é o número de pares ordenados no tempo na RM g .

Se o indicador de interesse for a diferença de duas razões, basta reescrever a equação (4.9), eliminando a parcela P_g^t e utilizar o mesmo estimador de variância (4.11) como é descrito na equação a seguir.

$$\begin{aligned} \hat{V}(\hat{R}_g^t - \hat{R}_g^{t-k}) \cong & \sum_{h=1}^{H_g} \sum_{i=1}^{n_{gh}^t} \frac{(z_{ghi}^t - \bar{z}_{gh}^t)^2}{n_{gh}^t - 1} + \\ & + \sum_{h=1}^{H_g} \sum_{i=1}^{n_{gh}^{t-k}} \frac{(z_{ghi}^{t-k} - \bar{z}_{gh}^{t-k})^2}{n_{gh}^{t-k} - 1} - 2 \sum_{h=1}^{H_g} \sum_{i=1}^{n_{gh}^*} \frac{(z_{ghi}^{t^*} - \bar{z}_{gh}^{t^*})(z_{ghi}^{(t-k)^*} - \bar{z}_{gh}^{(t-k)^*})}{n_{gh}^* - 1} \end{aligned} \quad (4.12)$$

5. Estimador da variância do estimador da diferença de totais e razões obtidos considerando pós-estratificação para o conjunto das seis Regiões metropolitanas.

Seja Y_C^t o total de uma quantidade de interesse para o conjunto das seis regiões metropolitanas da pesquisa no instante de tempo t . Seu estimador \hat{Y}_C^t não tendencioso é dado por.

$$\hat{Y}_C^t = \sum_{g=1}^6 \hat{Y}_g^t \frac{P_g^t}{\hat{P}_g^t} \quad (5.1)$$

Como foi visto na seção anterior, \hat{Y}_g^t corresponde ao estimador natural do desenho para o total de uma característica observada na pesquisa no instante de tempo t na RM g .

Analogamente o estimador desta quantidade de interesse no instante de tempo $t - k$ passa a ser representado através de:

$$\hat{Y}_C^{t-k} = \sum_{g=1}^6 \hat{Y}_g^{t-k} \frac{P_g^{t-k}}{\hat{P}_g^{t-k}} \quad (5.2)$$

O estimador da diferença de dois totais em dois instantes distintos de tempo para uma determinada variável, é dado por:

$$(\hat{Y}_C^t - \hat{Y}_C^{t-k}) = \sum_{g=1}^6 \hat{Y}_g^t \frac{P_g^t}{\hat{P}_g^t} - \sum_{g=1}^6 \hat{Y}_g^{t-k} \frac{P_g^{t-k}}{\hat{P}_g^{t-k}} \quad (5.3)$$

O estimador da variância do estimador da diferença entre dois totais pode ser obtido através da seguinte equação:

$$\hat{V}(\hat{Y}_C^t - \hat{Y}_C^{t-k}) = \sum_{g=1}^6 \hat{V}\left(\hat{Y}_g^t \frac{P_g^t}{\hat{P}_g^t}\right) + \sum_{g=1}^6 \hat{V}\left(\hat{Y}_g^{t-k} \frac{P_g^{t-k}}{\hat{P}_g^{t-k}}\right) - 2\left(\sum_{g=1}^6 \text{Cov}\left(\left(\hat{Y}_g^t \frac{P_g^t}{\hat{P}_g^t}\right)\left(\hat{Y}_g^{t-k} \frac{P_g^{t-k}}{\hat{P}_g^{t-k}}\right)\right)\right) \quad (5.4)$$

Seja R_C^t a razão entre dois totais da pesquisa, no instante de tempo t , o estimador desta quantidade é dado pela seguinte expressão:

$$\hat{R}_C^t = \frac{\left(\sum_{g=1}^6 \hat{Y}_g^t \frac{P_g^t}{\hat{P}_g^t}\right)}{\left(\sum_{g=1}^6 \hat{X}_g^t \frac{P_g^t}{\hat{P}_g^t}\right)} = \frac{\hat{Y}_C^t}{\hat{X}_C^t} \quad (5.5)$$

Analogamente o estimador desta quantidade de interesse no instante $t - k$ é obtido através da seguinte equação:

$$\hat{R}_C^{t-k} = \frac{\left(\sum_{g=1}^6 \hat{Y}_g^{t-k} \frac{P_g^{t-k}}{\hat{P}_g^{t-k}}\right)}{\left(\sum_{g=1}^6 \hat{X}_g^{t-k} \frac{P_g^{t-k}}{\hat{P}_g^{t-k}}\right)} = \frac{\hat{Y}_C^{t-k}}{\hat{X}_C^{t-k}} \quad (5.6)$$

O estimador da diferença entre estes dois estimadores de razão é dado por:

$$\left(\hat{R}_C^t - \hat{R}_C^{t-k}\right) = \left(\frac{\sum_{g=1}^6 \hat{Y}_g^t \frac{P_g^t}{\hat{P}_g^t}}{\sum_{g=1}^6 \hat{X}_g^t \frac{P_g^t}{\hat{P}_g^t}}\right) - \left(\frac{\sum_{g=1}^6 \hat{Y}_g^{t-k} \frac{P_g^{t-k}}{\hat{P}_g^{t-k}}}{\sum_{g=1}^6 \hat{X}_g^{t-k} \frac{P_g^{t-k}}{\hat{P}_g^{t-k}}}\right) = \frac{\hat{Y}_C^t}{\hat{X}_C^t} - \frac{\hat{Y}_C^{t-k}}{\hat{X}_C^{t-k}} \quad (5.7)$$

$$\text{Considerando } F(\hat{Y}_1^t \text{ K } \hat{Y}_6^t, \hat{X}_1^t \text{ K } \hat{X}_6^t, \hat{P}_1^t \text{ K } \hat{P}_6^t) = \frac{\sum_{g=1}^6 \hat{Y}_g^t \frac{P_g^t}{\hat{P}_g^t}}{\sum_{g=1}^6 \hat{X}_g^t \frac{P_g^t}{\hat{P}_g^t}} \quad (5.8)$$

Temos as seguintes derivadas parciais necessárias na construção da variável linearizada:

$$\frac{\partial F(\hat{Y}_1^t \text{ K } \hat{Y}_6^t, \hat{X}_1^t \text{ K } \hat{X}_6^t, \hat{P}_1^t \text{ K } \hat{P}_6^t)}{\partial \hat{Y}_g^t} = \frac{\frac{P_g^t}{\hat{P}_g^t}}{\sum_{g=1}^6 \hat{X}_g^t \frac{P_g^t}{\hat{P}_g^t}} \quad (5.9)$$

$$\frac{\partial F(\hat{Y}_1^t \text{ K } \hat{Y}_6^t, \hat{X}_1^t \text{ K } \hat{X}_6^t, \hat{P}_1^t \text{ K } \hat{P}_6^t)}{\partial \hat{X}_g^t} = - \frac{\sum_{g=1}^6 \hat{Y}_g^t \frac{P_g^t}{\hat{P}_g^t}}{\left(\sum_{g=1}^6 \hat{X}_g^t \frac{P_g^t}{\hat{P}_g^t} \right)^2} \left(\frac{P_g^t}{\hat{P}_g^t} \right) \quad (5.10)$$

$$\frac{\partial F(\hat{Y}_1^t \text{ K } \hat{Y}_6^t, \hat{X}_1^t \text{ K } \hat{X}_6^t, \hat{P}_1^t \text{ K } \hat{P}_6^t)}{\partial \hat{P}_g^t} = \frac{- \left(\sum_{g=1}^6 \hat{X}_g^t \frac{P_g^t}{\hat{P}_g^t} \right) \hat{Y}_g^t \frac{P_g^t}{(\hat{P}_g^t)^2} + \left(\sum_{g=1}^6 \hat{Y}_g^t \frac{P_g^t}{\hat{P}_g^t} \right) \hat{X}_g^t \frac{P_g^t}{(\hat{P}_g^t)^2}}{\left(\sum_{g=1}^6 \hat{X}_g^t \frac{P_g^t}{\hat{P}_g^t} \right)^2} \quad (5.11)$$

Portanto, neste caso, z_{ghijk}^t corresponde ao valor da variável linearizada Z da k-ésima observação no j-ésimo domicílio da i-ésima unidade primária de amostragem no estrato de seleção h, na g-ésima Região Metropolitana no instante de tempo t.

$$z_{ghijk}^t = \left\{ \frac{\frac{P_g^t}{\hat{P}_g^t} y_{ghijk}^t - \frac{\sum_{g=1}^6 \hat{Y}_g^t \frac{P_g^t}{\hat{P}_g^t}}{\left(\sum_{g=1}^6 \hat{X}_g^t \frac{P_g^t}{\hat{P}_g^t} \right)^2} \left(\frac{P_g^t}{\hat{P}_g^t} \right) x_{ghijk}^t + \frac{\left(\sum_{g=1}^6 \hat{Y}_g^t \frac{P_g^t}{\hat{P}_g^t} \right) \hat{X}_g^t \frac{P_g^t}{(\hat{P}_g^t)^2} - \left(\sum_{g=1}^6 \hat{X}_g^t \frac{P_g^t}{\hat{P}_g^t} \right) \hat{Y}_g^t \frac{P_g^t}{(\hat{P}_g^t)^2}}{\left(\sum_{g=1}^6 \hat{X}_g^t \frac{P_g^t}{\hat{P}_g^t} \right)^2} P_{ghijk}^t \right\} w_{ghijk}^t \quad (5.12)$$

Para melhor visualização a expressão anterior pode ser representada da seguinte maneira:

$$z_{ghijk}^t = \frac{P_g^t}{\hat{P}_g^t \hat{X}_C^t} \left[y_{ghijk}^t - R_C x_{ghijk}^t + \left(\hat{Y}_C^t \frac{\hat{X}_g^t}{\hat{P}_g^t} - \hat{X}_C^t \frac{\hat{Y}_g^t}{\hat{P}_g^t} \right) \frac{P_{ghijk}^t}{\hat{X}_C^t} \right] w_{ghijk}^t \quad (5.13)$$

Portanto o estimador da variância do estimador da diferença de dois estimadores de razão passa a ser descrito através da seguinte aproximação:

$$\hat{V}(R_C^t - R_C^{t-k}) \cong \sum_{g=1}^6 \sum_{h=1}^{H_g} \sum_{i=1}^{n_{gh}^t} \frac{(z_{ghi}^t - \bar{z}_{gh}^t)^2}{n_{gh}^t - 1} + \sum_{g=1}^6 \sum_{h=1}^{H_g} \sum_{i=1}^{n_{gh}^{t-k}} \frac{(z_{ghi}^{t-k} - \bar{z}_{gh}^{t-k})^2}{n_{gh}^{t-k} - 1} - 2 \sum_{g=1}^6 \sum_{h=1}^{H_g} \sum_{i=1}^{n_{gh}^*} \frac{(z_{ghi}^{t*} - \bar{z}_{gh}^{t*})(z_{ghi}^{(t-k)*} - \bar{z}_{gh}^{(t-k)*})}{n_{gh}^* - 1} \quad (5.14)$$

6. Intervalos de confiança.

Seja q um parâmetro de interesse e \hat{q} seu estimador amostral não tendencioso.

Logo:

$$E_p(\hat{q}) = q \quad (6.1)$$

A variância do estimador de q sobre o plano amostral adotado, é representada por:

$$V_p(\hat{q}) = s_p^2 \text{ e seu estimador, } \hat{V}_p(\hat{q}) = \hat{s}_p^2 \quad (6.2)$$

Ao calcularmos um intervalo de confiança para q utilizando \hat{q} como estimador, devemos sempre nos basear em estimativas de variância de \hat{q} que considerem o método de seleção da amostra, pois ao ignorar o processo de seleção no cálculo da estimativa da variância do estimador \hat{q} , estamos alterando o nível de significância do intervalo, o que pode levar a conclusões equivocadas.

Portanto, o intervalo de confiança do parâmetro q é dado pela seguinte expressão:

$$P\left(\hat{q} - z_{\frac{\alpha}{2}} \cdot \sqrt{\hat{s}_p^2} \leq q \leq \hat{q} + z_{\frac{\alpha}{2}} \cdot \sqrt{\hat{s}_p^2}\right) = 1 - \alpha \quad (6.3)$$

7. Aplicação.

Como foi escrito no capítulo introdutório foram calculados intervalos de confiança para um conjunto de 41 variáveis da pesquisa, objetivando validar a existência de diferenças significativas entre as estimativas em pares de meses.

Primeiramente foram calculadas estimativas pontuais de cada variável para o instante de tempo t e em seguida para $t - k$. Onde k assume os valores 1, 3, 6. De posse do valor das estimativas foram calculadas as diferenças entre os dois instantes de tempo.

Depois foram estimadas as variâncias destas diferenças, utilizando o Método de Linearização de Taylor, descrito na seção 3 e finalmente foram calculados os intervalos de confiança para as diferenças considerando o nível de significância $\alpha = 0,05$.

O critério para validação das diferenças nos indicadores foi verificar se o valor zero estava contido no intervalo. Caso fosse verificado a existência destes valores nos intervalos, a conclusão seria de que não existe diferença significativa entre os valores do indicador para os instantes de tempo considerados.

Os resultados foram estimados para o período de janeiro de 2002 a abril de 2003 tanto para valores referentes as Regiões Metropolitanas como para o conjunto das seis.

Os 41 indicadores considerados no estudo foram os seguintes:

- Total de pessoas em idade ativa (01.1)²
- Total de pessoas economicamente ativas (02.1);
- Total de pessoas ocupadas (03.1);
- Total de empregados (03.05.1);
- Total de empregados com carteira de trabalho assinada (03.05.1.3);
- Total de empregados sem carteira de trabalho assinada (03.05.1.4);
- Total de trabalhadores por conta própria (03.05.2);
- Total de empregadores (03.05.3);
- Total de trabalhadores não remunerados, conta própria ou empregador (03.05.4);
- Total de trabalhadores não remunerados, empregados (03.05.5);
- Total de pessoas ocupadas na indústria extrativa e de produção, e produção e distribuição de eletricidade, gás, e água (03.06.1);
- Total de pessoas ocupadas na construção (03.06.2);

² Corresponde ao código do indicador.

- Total de pessoas ocupadas no comércio, reparação de veículos automotores e de objetos pessoais e domésticos e comércio a varejo de combustíveis etc (03.06.3);
- Total de pessoas ocupadas no serviços prestados à empresa, aluguéis, atividades imobiliárias e intermediação financeira (03.06.4);
- Total de pessoas ocupadas na administração pública, defesa, seguridade social, educação, saúde e serviços sociais (03.06.5);
- Total de pessoas ocupadas em serviços domésticos (03.06.6);
- Total de pessoas ocupadas em outros serviços (03.06.7);
- Total de pessoas ocupadas em outras atividades (03.06.8);
- Total de empregados no setor privado (03.07.1);
- Total de empregados com carteira de trabalho assinada no setor privado (03.07.1.1);
- Total de desocupados (04.1);;
- Total de pessoas não economicamente ativas (05.1);
- Total de pessoas não economicamente ativas que não gostariam de trabalhar (05.5.0);
- Total de pessoas não economicamente ativas que gostariam e estavam disponíveis para trabalhar (05.5.1);
- Total de pessoas não economicamente ativas que gostariam e não estavam disponíveis para trabalhar (05.5.2);
- Total de pessoas não economicamente ativas marginalmente ligadas à PEA (05.6);
- Total de pessoas não economicamente ativas marginalmente ligadas à PEA desalentadas por vários motivos (05.7.1);

- Total de pessoas não economicamente ativas marginalmente ligadas à PEA desalentadas por não conseguir qualquer trabalho (05.7.2);
- Taxa de atividade (06.3.1);
- Razão Total de pessoas ocupadas/Total de pessoas economicamente ativas (06.3.2);
- Razão Total de pessoas desocupadas com tempo de procura de 30 dias/Total de pessoas em idade ativa (06.3.3);
- Taxa de inatividade (06.3.4);
- Taxa de ocupação (07.3);
- Taxa de desocupação (07.4);
- Taxa de desocupação - sexo masculino (21.1.1);
- Taxa de desocupação - sexo feminino (21.1.2);
- Razão Total de pessoas com rendimento por hora efetivos inferior ao salário mínimo por hora/ Total de pessoas ocupadas (26);
- Total de pessoas que saíram do último trabalho no período de 365 dias (32);
- Total de pessoas subocupadas na semana de referência por insuficiência de horas trabalhadas (33);
- Total de pessoas com rendimento efetivo por hora inferior ao salário mínimo por hora (34);
- Razão Total de pessoas subocupadas na semana de referência por insuficiência horas trabalhadas/Total de pessoas ocupadas (35);

Cabe ressaltar que os resultados referentes ao agregado das seis regiões devem ser vistos com certa cautela quando é feita a comparação com os meses de janeiro e

fevereiro de 2002 onde somente há resultados para as Regiões Metropolitanas de Recife, Belo Horizonte, Rio de Janeiro e São Paulo.

7.1 Resumo dos resultados

As tabelas contidas no Anexo 1 ilustram as estimativas pontuais dos indicadores citados anteriormente, quais diferenças entre pares de meses foram consideradas significativas para cada indicador, bem como os limites de confiança para construção dos intervalos referentes as diferenças temporais.

Além destas tabelas, os resultados podem ser interpretados através da análise gráfica como mostram os gráficos contidos no Anexo 2.

Os gráficos 1 a 9 mostram os resultados para as diferenças $[t - (t - 1)]$, $[t - (t - 3)]$ e $[t - (t - 6)]$ para três indicadores da Pesquisa Mensal de Emprego.

Os indicadores foram escolhidos de maneira que teríamos uma análise para os estimadores descritos nas seções 4 e 5.

De acordo com o Gráfico 1 somente foi verificada diferença significativa de um mês em relação ao mês anterior no que diz respeito ao Total de Empregados de agosto de 2002 em relação a julho de 2002, indicando uma diferença positiva o que indica um crescimento significativo do Total de Empregados neste período.

Os Gráficos 2 e 3 mostram que diferenças significativas acerca destas estimativas ocorrem com mais frequência quando a comparação é feita em relação a períodos mais longos.

Através do gráfico 4 pode-se observar poucas variações significativas no valor da Taxa de Desocupação em relação ao mês anterior, com indicação de decréscimo significativo para a taxa em maio de 2002 em relação a abril de 2002. Entretanto, também foram verificados crescimentos significativos da taxa em janeiro e março de 2003.

Observando o comportamento dos gráficos 5 e 6 verifica-se que ao se comparar resultados através períodos mas extensos verifica-se tendências mais fortes acerca dos indicadores.

Os resultados encontrados no Gráfico 7 mostram a evolução das diferenças mensais para o total de pessoas subocupadas na semana de referência por insuficiência de horas trabalhadas. Foram encontradas diferenças significativas nos meses de janeiro de 2002, janeiro e fevereiro de 2003 em relação aos meses anteriores.

E como nos dois indicadores analisados anteriormente, observa-se que a quantidade de meses com diferenças significativas aumenta quando a comparação é feita com um mês mais distante, isto pode ser visto nos gráficos 8 e 9.

8. Bibliografia.

Cochran, W.G. Sampling Techniques(3rd ed.). New York, Wiley, 1977.

Corrêa, S.T., Freitas, M.P.S. e P.L. Nascimento Silva. Estimaco de Varincia para o Estimador da Diferena entre Duas Taxas na Pesquisa Mensal de Emprego. IBGE, Rio de Janeiro, 2001.

Hansen, M. H., Hurvitz, W. N., Madow, W. G. Sample Survey Methods and Theory, Vol I e II. New York, Wiley, 1953.

Pesquisa Mensal de Emprego, Srie Relatrios Metodolgicos, IBGE,2003.

Smdal, C.E., B. Sewnsson e J.H. Wretman. Model Assisted Survey Sampling. Springer-Verlag, Nova Iorque, 1992.

Shah, B.V., B.G. Barnwell e G.S. Biele. SUDAAN User's Manual, release 7.5. Research Triangle Park, NC: Research Triangle Institute, 1997.

Shah, B.V., R.E. Folsom, L.M. LaVange, S.C. Wheelless, K.E. Boyle e R.L. Williams - Statistical Methods and Mathematical Algorithms Used in SUDAAN. Research Triangle Park, NC: Research Triangle Institute, 1995.

Wolter, K.M. Introduction to Variance Estimation. New York: Springer-Verlag,1985. 427p.