

UNIVERSIDADE FEDERAL DE JUIZ DE FORA

INSTITUTO DE CIÊNCIAS EXATAS

DEPARTAMENTO DE ESTATÍSTICA

Daniel Gustavo Moreira

Um Estudo Sobre o Efeito do Plano Amostral da PNAD Contínua

Juiz de Fora

2019

Daniel Gustavo Moreira

Um Estudo Sobre o Efeito do Plano Amostral da PNAD Contínua

Monografia apresentada ao Curso de Estatística da Universidade Federal de Juiz de Fora, como requisito parcial para a obtenção do título de Bacharel em Estatística.

Orientador: Marcel de Toledo Vieira

Coorientador: Ricardo da Silva Freguglia

Juiz de Fora

2019

Ficha catalográfica elaborada através do programa de geração automática da Biblioteca Universitária da UFJF, com os dados fornecidos pelo(a) autor(a)

Moreira, Daniel Gustavo.
Um Estudo Sobre o Efeito do Plano Amostral da PNAD Contínua / Daniel Gustavo Moreira. -- 2019.
67 p.

Orientador: Marcel de Toledo Vieira
Coorientador: Ricardo da Silva Freguglia
Trabalho de Conclusão de Curso (graduação) - Universidade Federal de Juiz de Fora, Instituto de Ciências Exatas, 2019.

1. Amostragem complexa. 2. EPA. 3. Mercado de trabalho. 4. Análise de dados amostrais. I. Vieira, Marcel de Toledo , orient. II. Freguglia, Ricardo da Silva , coorient. III. Título.

Daniel Gustavo Moreira

Um Estudo Sobre o Efeito do Plano Amostral da PNAD Contínua

Monografia apresentada ao Curso de Estatística da Universidade Federal de Juiz de Fora, como requisito parcial para a obtenção do título de Bacharel em Estatística.

Aprovada em 28 de junho de 2019.

BANCA EXAMINADORA

Prof. Dr. Marcel de Toledo Vieira - Orientador
Universidade Federal de Juiz de Fora

Prof. Dr. Ricardo da Silva Freguglia - Coorientador
Universidade Federal de Juiz de Fora

Prof. Dr. Augusto Carvalho Souza
Universidade Federal de Juiz de Fora

Prof. Dr. Tiago Maia Magalhães
Universidade Federal de Juiz de Fora

AGRADECIMENTOS

Agradeço e dedico essa monografia às seguintes pessoas:

À minha mãe Maria e ao meu pai Antônio, por serem pessoas fortes, de caráter e que sempre me apoiaram.

Ao meu irmão Patrick, por ser um homem de bem, de valores únicos e pela eterna amizade que temos.

À Izabel, minha amada esposa, que está sempre ao meu lado lutando minhas batalhas com fibra e garra, além de ser a luz da minha vida.

À tia Sandra, por ser uma pessoa doce e de paz.

Ao professor Marcel e Ricardo, que me orientaram com tanto saber e dedicação, e ainda, dispensaram boas horas de suas paciências comigo.

Aos meus amigos, que estão pelo mundo, mas quando nos encontramos é como se o tempo não tivesse passado.

Aos amigos do curso de Estatística que entendem a dureza da graduação.

Aos professores do curso, por terem me conduzido com tanto empenho por esse caminho de conhecimento.

À Faculdade de Economia, por ter permitido a utilização do Software *Stata 15* e em especial ao Laboratório de Estudo Econômicos (ECONS), onde tive a oportunidade de ser bolsista e também germinou a ideia para o tema dessa Monografia.

Ao Instituto Brasileiro de Geografia e Estatística (IBGE), por ter disponibilizado a base de dados da Pesquisa Nacional por Amostra de Domicílios Contínua (PNAD Contínua).

E por fim, à Universidade Federal de Juiz de Fora (UFJF), por todo o suporte dando durante estes anos.

“There is no knowledge that is not power”

Ralph Waldo Emerson

RESUMO

No Brasil são realizadas inúmeras pesquisas para estimação de vários indicadores sociais e econômicos. Alguns destes estudos são complexos e de grande abrangência, o que resulta em um grande volume de dados produzidos. Logo, são necessários métodos de amostragem que consigam levar em consideração todos os elementos que compõem a população em uma amostra representativa e assim construir estimações. Com isso, a amostragem se faz presente desde a coleta de dados até a análise dos mesmos. Para realizarmos este trabalho usamos dados da Pesquisa Nacional por Amostra de Domicílios Contínua (PNAD-C) realizada pelo Instituto Brasileiro de Geografia e Estatística (IBGE) desde 2012. Temos por objetivo verificar os problemas ocasionados por ignorar o plano amostral da PNAD Contínua na análise de dados. Iremos, de forma resumida e introdutória, apresentar alguns planos amostrais mais usuais e métodos para análise de dados amostrais complexos. Também apresentaremos resultados de taxas que foram estimadas, a fim de ilustrar o objetivo do trabalho. Para isso foi usado o software *Stata 15* que possibilitou os cálculos das estimativas, intervalos de confiança, erros-padrão e EPAs, considerando e não considerando o plano amostral da PNAD-C. Observamos que ao ignorarmos o plano amostral ocorreu uma diferença nos valores obtidos para as estimações das médias se compararmos com as médias estimadas obtidas sob o plano amostral da pesquisa e ainda ocorreu a subestimação dos erros que afetou diretamente os intervalos de confiança produzidos.

Palavras-chave: Amostragem complexa; EPA; Mercado de trabalho; Análise de dados amostrais.

ABSTRACT

In Brazil, numerous surveys are conducted to estimate various social and economic indicators. Some of these studies are complex and wide ranging, resulting in a large volume of data produced. Therefore, sampling methods are necessary to take into account all the elements that make up the population in a representative sample and thus to construct estimates. With this, the sampling takes place from the collection of data to the analysis of the same. In order to carry out this work we use data from the PNAD-C conducted by the IBGE since 2012. We aim to verify the problems caused by ignoring the sampling plan of the Continuous PNAD in the analysis of data. We will summarize and introduce some more usual sampling plans and methods for analyzing complex sample data. We will also present results of rates that were estimated in order to illustrate the purpose of the work. Stata 15 software was used to calculate estimates, confidence intervals, standard errors and EPAs, considering and not considering the sample plan of the PNAD-C. We observed that when we ignored the sampling plan, there was a difference in the values obtained for the estimates of the means if we compare with the estimated means obtained under the sampling plan of the research and also the underestimation of the errors occurred that directly affected the confidence intervals produced.

Keywords: Complex sampling; EPA; Job market; Analysis of sample data.

SUMÁRIO

1	INTRODUÇÃO	8
2	PNAD CONTÍNUA	9
2.1	A PNAD CONTÍNUA	9
2.2	PLANO AMOSTRAL DA PNAD CONTÍNUA	12
3	MÉTODOS DE AMOSTRAGEM	13
3.1	AMOSTRAGEM ESTRATIFICADA	13
3.1.1	Notações importantes	14
3.1.2	Amostragem Estratificada Simples (AES)	15
3.1.3	Alocação da amostra pelos estratos	16
3.2	AMOSTRAGEM POR CONGLOMERADOS	18
3.2.1	Amostragem por Conglomerados em Um Estágio	19
3.2.2	Notações importantes	20
3.2.3	Amostragem Conglomerada Simples (ACS)	20
3.2.4	Amostragem por Conglomerados em Dois Estágios	21
3.2.5	Amostragem por Conglomerados em Dois Estágios com Amostragem Aleatória Simples (AC2 – AASc)	22
3.3	AMOSTRAGEM COM PROBABILIDADES PROPORCIONAIS AO TAMANHO	23
3.3.1	Estimador Horvitz-Thompson (HT)	24
3.4	ESTIMADOR RAZÃO	25
3.4.1	Estimador Razão aplicado pelo Stata	26
3.5	EFEITO DO PLANO AMOSTRAL (EPA)	28
3.6	TAXAS	28
4	EPA NA PNAD CONTÍNUA	31
4.1	BANCO DE DADOS	31
4.2	VARIÁVEIS	32
4.3	RESULTADOS	37
5	CONSIDERAÇÕES FINAIS	45
	REFERÊNCIAS	47
	APÊNDICE	51
	APÊNDICE A – CÓDIGO UTILIZADOS	51

1 – INTRODUÇÃO

Atualmente no Brasil existem várias pesquisas que visam obter informações sobre várias características da população, tais como: sexo, idade, educação, trabalho e rendimento, domicílios, migração, fecundidade, nupcialidade, etc. Podendo ter periodicidade variável ou não e com temas específicos que investigam aspectos demográficos, sociais e econômicos.

Em pesquisas de grande escala, que abrangem o país, chama a atenção a Pesquisa Nacional por Amostras de Domicílios Contínua (PNADC, 2019), dada sua capacidade de captar uma grande quantidade de informação de cada unidade domiciliar pesquisada e gerar inúmeros indicadores sobre o panorama socioeconômico brasileiro. Então, por ser uma pesquisa de abrangência nacional e ainda, ter divulgação trimestral de informações sobre indicadores relacionados à força de trabalho, devemos nos questionar se há problemas em ignorar o método de amostragem adotado no processo de produção das estimativas.

Tendo em mente a importância da PNADC, realizamos um estudo para verificar os problemas ocasionados por ignorar o plano amostral da pesquisa. Para apurar os possíveis problemas iremos verificar a eficiência do plano amostral adotado na pesquisa em relação a um plano amostral padrão, sob AAS, através do Efeito do Plano de Amostral (EPA), que é uma razão entre variâncias de um estimador, comparando dois planos amostrais.

De forma breve, a Pesquisa Nacional por Amostras de Domicílios Contínua pode ser definida como: uma pesquisa que visa produzir indicadores para acompanhar as flutuações trimestrais e a evolução, a médio e longo prazos, da força de trabalho e outras informações necessárias para o estudo e desenvolvimento socioeconômico do País.

Além disso, é crucial um entendimento sobre a base de dados extremamente complexa e minuciosa e de seus questionamentos aos pesquisados. Visto o volume de variáveis criadas, que propiciam uma série de interpretações sobre vários aspectos da população brasileira para compreender o mercado de trabalho.

Nos capítulos seguintes iremos apresentar melhor a abrangência da pesquisa no Brasil, como são dadas as áreas censitárias de interesse a serem pesquisadas e sua periodicidade de aplicação, além da metodologia amostral utilizada com suas características particulares.

Também definiremos alguns métodos de amostragem - Amostragem Estatística, Amostragem por Conglomerados e Amostragem por PPT - que irão contribuir para o entendimento do trabalho proposto e alguns indicadores como: taxa de ocupação, taxa de desemprego, nível de ocupação, nível de desocupação, entre outras.

No Capítulo 2 apresentaremos a PNAD Contínua, mostrando como foi a sua formação, implementação, objetivos da pesquisa e a forma de elaboração do plano amostral. O Capítulo 3 aborda, de forma resumida, métodos de amostragem e inferência, além dos conceitos das taxas utilizadas no estudo. Na sequência, o Capítulo 4 aborda definições do banco de dados, software utilizado e os resultados obtidos. No Capítulo 5, será dada a conclusão final. Por fim, a seção Referências, onde estão as fontes consultadas para a Monografia e a seção Apêndice, onde estão os códigos criados para analisar o banco de dados da PNAD Contínua.

2 – PNAD CONTÍNUA

2.1. A PNAD CONTÍNUA

O IBGE com intuito de aprimorar seu sistema de levantamento de informações e atender o constante questionamento por parte de órgãos governamentais, centros de estudo, pesquisadores e organismos internacionais, reformulou suas pesquisas por amostragem de domicílios, através da implantação do Sistema Integrado de Pesquisas Domiciliares (SIPD) (IBGE, 2007), do qual PNAD Contínua é um dos pilares.

A PNAD Contínua, visando produzir indicadores conjunturais relativos à força de trabalho substituiu as estatísticas sobre mercado de trabalho obtidas pela Pesquisa Mensal de Emprego (PME) e com isso aumentou a cobertura para o território

nacional. Também substituiu a Pesquisa Nacional por Amostra de Domicílios (PNAD), o que permitiu a divulgação periódica das informações sobre trabalho e ainda a análise conjuntural do tema. Incorporou também atualizações conceituais referentes ao tema trabalho, tendo como base recomendações internacionais vigentes no momento da concepção da pesquisa, nas discussões preparatórias para a 19ª Conferência Internacional de Estatísticos do Trabalho (CIET, 2013), da OIT, e em recomendações adotadas pela Conferência em outubro de 2013 sobre trabalho, ocupação e subutilização da mão de obra.

A PNAD Contínua foi implantada, em caráter experimental em outubro de 2011, tendo o principal objetivo ajustar os processos referentes à pesquisa (IBGE, 2014 p. 6). E as seguintes áreas receberam a pesquisa de forma experimental:

a) 20 Regiões Metropolitanas: Manaus, Belém, Macapá, São Luís, Fortaleza, Natal, João Pessoa, Recife, Maceió, Aracaju, Salvador, Belo Horizonte, Vitória, Rio de Janeiro, São Paulo, Curitiba, Florianópolis, Porto Alegre, Vale do Rio Cuiabá e Goiânia;

b) na Região Integrada de Desenvolvimento (RIDE) da Grande Teresina;

c) em 5 capitais: Porto Velho, Rio Branco, Boa Vista, Palmas e Campo Grande; e

d) no Distrito Federal

Em janeiro de 2012 passou a ser implantada em todo o território nacional, fazendo parte em definitivo do conjunto de pesquisas do IBGE.

A pesquisa é realizada por meio de uma amostra probabilística de domicílios, onde há a extração de uma amostra mestra (QUINTSLR et al, 2007, p. 15) de setores censitários, assim garantindo a representatividade dos resultados para os diversos níveis geográficos como: Brasil, Grandes Regiões, Unidades da Federação e Regiões Metropolitanas que incluem os municípios das capitais. A cada trimestre, a PNAD Contínua investiga aproximadamente 211.000 domicílios em 16.000 setores censitários. Números superiores a PNAD, logo permite um ganho considerável na precisão das estimativas, principalmente nas Unidades de Federação de menor tamanho de população e nas áreas rurais (IBGE, 2014, p. 7).

As informações sobre o tema trabalho são captadas pela PNAD Contínua em dois questionários (IBGE, 2014, p. 8):

- Reduzido, com informações sobre o trabalho que gera rendimentos para o domicílio, em que a produção é voltada para o mercado. As informações são investigadas em 100% dos domicílios pesquisados, além das informações básicas dos moradores (condição no domicílio, sexo, idade, cor ou raça e educação) de forma trimestral; e

- Ampliado, aplicado na primeira entrevista de cada domicílio, que inclui variáveis contidas no reduzido e informações sobre outras formas de trabalho, cuja produção não é voltada para o mercado. Abrangendo informações sobre outras formas de trabalho como a produção para o próprio consumo, o trabalho voluntário, o cuidado de pessoas do domicílio e os afazeres domésticos. Esse questionário é investigado em 20% da amostra de cada trimestre e os seus dados são acumulados e produz resultados anuais referentes a um ano civil.

Os temas trabalho, rendimento e as características básicas dos moradores foram os primeiros a serem implantados na pesquisa. Outros temas permanentes da pesquisa - migração, fecundidade e trabalho infantil - poderão ser investigados em um trimestre específico ou aplicados em uma parte da amostra a cada trimestre gerando um acumulado e produzindo resultados anuais. Os temas suplementares com período variável, também poderão ser investigados em um trimestre ou acumulados ao longo de um ano.

Alguns indicadores básicos relativos à análise do mercado de trabalho: pessoas em idade de trabalhar, força de trabalho na semana de referência, condição de ocupação na semana de referência, posição na ocupação do trabalho principal, categoria do emprego do trabalho principal, taxa de participação na força de trabalho, nível da ocupação, nível da desocupação, taxa de desocupação, taxa de desocupação, taxa de formalidade e taxa de informalidade.

2.2. PLANO AMOSTRAL DA PNAD CONTÍNUA

A Pesquisa Nacional por Amostras de Domicílios Contínua (PNADC) abrange todo o território nacional e é realizada por meio de uma amostra probabilística de domicílios, extraída de uma amostra mestra de setores censitários, da Base Operacional Geográfica de 2010, exceto as áreas com características especiais como: aldeias indígenas, quartéis, bases militares, alojamentos, acampamentos, embarcações, barcos, navios, penitenciárias, colônias penais, presídios, cadeias, asilos, orfanatos, conventos, hospitais e agrovilas de projetos de assentamentos rurais, e também os setores censitários localizados em Terras Indígenas. Todas as pessoas moradoras em domicílios particulares permanentes da área de abrangência compõem a população alvo da pesquisa.

A PNAD Contínua adota o plano amostral conglomerado em dois estágios de seleção com estratificação das unidades primárias de amostragem (UPA) (IBGE, p. 21, 2014) e cada UPA foi definida considerando os setores censitários devendo possuir ao menos 60 domicílios particulares permanentes (DPPs), incluindo os ocupados, os ocupados sem entrevista realizada e os vagos. Para setores censitários, nos quais os tamanhos eram inferiores a 60 (DPPs), houve o agrupamento no mesmo subdistrito, respeitando a contiguidade, o tipo e a situação dos setores até ocorrer a formação de 60 DPPs, o mínimo estabelecido para formar uma UPA.

- Primeiro estágio: são selecionadas UPAs com probabilidade proporcional ao número de domicílios dentro de cada estrato e a estratificação adotada é a definida para todo o SIPD (QUINTSLR, 2007). A seleção é feita do Cadastro Mestre, que contém para cada UPA informações sobre a dependência administrativa e algumas características sociodemográficas. Amostra Mestra para um trimestre é constituída pelas UPAs que compõem a amostra da PNAD Contínua.

- Segundo estágio: são selecionados 14 domicílios particulares permanentes ocupados dentro de cada UPA da amostra, por amostragem aleatória simples (AAS) (COCHRAN, 1977) do Cadastro Nacional de Endereços para Fins Estatísticos (CNEFE) (IBGE, 2013) atualizado.

A amostra de uma UPA e de domicílios é dividida pelos 3 meses de um trimestre, seguindo um esquema de rotação (IBGE, 2014 p,22), ou seja, periodicidade de

coleta trimestral, assim ao final deste ciclo serão produzidas as estimativas dos indicadores de interesse. O esquema de rotação da amostra adotado foi o esquema 1-2(5). Neste esquema o domicílio é entrevistado 1 mês e sai da amostra por 2 meses seguidos, sendo esta sequência repetida 5 vezes. São definidos 15 grupos de rotação de domicílios, dividindo a amostra de UPAs nestes grupos. A cada mês do trimestre são pesquisados 5 grupos, e ao final do período a amostra é acumulada para produção dos indicadores.

Além disso, o esquema 1-2(5) é o mais eficiente quando um dos principais interesses da pesquisa é a inferência a respeito de mudanças em indicadores trimestrais. Esse método de seleção da amostra dá a possibilidade de atualizar e ampliar o cadastro de seleção. Além de permitir acompanhar a evolução do crescimento das UPAs, mudanças na base operacional geográfica e renovação controlada da amostra.

3 – MÉTODOS DE AMOSTRAGEM

3.1. AMOSTRAGEM ESTRATIFICADA

Amostragem estratificada é uma técnica que pertence à família de amostras probabilísticas que consiste na divisão de uma população em grupos (estratos) de maneira que um indivíduo pode fazer parte apenas de um único estrato ou camada, dada algumas características conhecidas na população em estudo, sendo mutuamente exclusivos e exaustivos.

Para cada um dos estratos são selecionadas amostras em proporções convenientes e de forma independente. Também podemos usar outras técnicas de amostragem em cada estrato (amostra sistemática, aleatória com reposição ou sem reposição, etc).

Se há uma população muito heterogênea, temos razões custos que limitam o tamanho da amostra e torna-se impossível definir uma AASc da população toda com precisão razoável. Portanto, uma saída é o uso da amostragem estratificada, que reduz o erro amostral, dado que subdivide uma população em subpopulações internamente

mais homogêneas, formando grupos com variâncias menores que diminuiriam o erro amostral global.

quanto mais homogêneos os subgrupos, maior será a eficiência do plano amostral. Assim melhorando a precisão dos resultados ao realizar um estudo sobre a amostra em questão.

De acordo com Bolfarine e Bussab (2004, p.96).

[...] “O resultado será mais eficaz quanto maior for a habilidade do pesquisador em produzir estratos homogêneos”.

Motivos para estratificar:

- Obtendo um grau de homogeneidade dentro de cada estrato e um grau de heterogeneidade entre os estratos, podemos realizar estimações tanto para a população quanto para cada estrato;
- Resolvendo problemas administrativos, como o particionando de custos entre as entidades responsáveis pela pesquisa em cada estrato da pesquisa;
- Permitindo a adoção de procedimentos alternativos de estimação para algumas subpopulações. Considerado cada subpopulação específica em um estrato; e
- A estratificação pode gerar um ganho de precisão nas estimativas das características de toda a população.

3.1.1 Notações importantes

Aqui iremos expressar algumas notações utilizadas para o método de amostragem estratificada e indicando as unidades populacionais com os seguintes rótulos:

Definiremos $h = 1, \dots, H$ para indicar o estrato a que pertencem e $i = 1, \dots, N_h$ para indicar a unidade dentro do estrato.

O tamanho populacional como $N = N_1 + N_2 + \dots + N_H$ e o tamanho amostral como $n = n_1 + n_2 + \dots + n_H$.

Para o estrato h o total é dado da seguinte forma,

$$Y_h = \sum_{i=1}^{N_h} y_{hi},$$

a média como

$$\bar{Y}_h = \frac{Y_h}{N_h},$$

e a variância por

$$S_h^2 = \frac{\sum_{i=1}^{N_h} (y_{hi} - \bar{Y}_h)^2}{N_h - 1}.$$

E para a população o total é dado por

$$Y = \sum_{h=1}^H Y = \sum_{h=1}^H N_h \bar{Y}_h,$$

a média populacional

$$\bar{Y} = \frac{Y}{N} = \frac{\sum_{h=1}^H N_h \bar{Y}_h}{N} = \sum_{h=1}^H W_h \bar{Y}_h,$$

onde $W_h = \frac{N_h}{N}$,

e a variância populacional é

$$S^2 = \frac{\sum_{h=1}^H [(N_h - 1)S_h^2 + N_h(\bar{Y}_h - \bar{Y})^2]}{(N - 1)}.$$

3.1.2. Amostragem Estratificada Simples (AES)

É um caso especial da amostragem estratificada em que AAS (BOLFARINE; BUSSAB, 2004) é utilizada em todos os estratos, dado que, o tamanho n_h é conhecido de cada estrato.

Nos estratos temos os seguintes parâmetros:

O total no estrato, $\hat{Y}_h = \frac{N_h}{n_h} \sum_{i=1}^{n_h} y_{hi} = N_h \bar{y}_h = \sum_{i=1}^{n_h} w_h y_{hi}$, onde $w_h = \frac{N_h}{n_h}$.

A média é dada por $\bar{y}_h = \frac{\sum_{i=1}^{n_h} y_{hi}}{N_h}$, onde $E_{AAS}(\bar{y}_h) = \bar{Y}_h$; e

A variância por $s_h^2 = \sum_{i=1}^{n_h} \frac{(y_{hi} - \bar{y}_h)^2}{(n_h - 1)}$ e $E_{AAS}(s_h^2) = S_h^2$.

O total populacional, $\hat{Y}_{AES} = \sum_{h=1}^H \hat{Y}_h = \sum_{h=1}^H N_h \bar{y}_h$;

A média populacional, $\bar{y}_{AES} = \sum_{h=1}^H W_h \bar{y}_h$, em que, $W_h = \frac{N_h}{N}$.

Propriedades de \bar{y}_{AES} (I),

$$E_{AES}(\bar{y}_{AES}) = \bar{Y} \text{ (não viciado)}$$

Propriedades de \bar{y}_{AES} (II),

$$V_{AES}(\bar{y}_{AES}) = \sum_{h=1}^H W_h^2 \left(1 - \frac{n_h}{N_h}\right) \frac{S_h^2}{n_h} = \sum_{h=1}^H W_h^2 \left(\frac{1}{n_h} - \frac{1}{N_h}\right) S_h^2$$

e

$$\hat{V}_{AES}(\bar{y}_{AES}) = \sum_{h=1}^H W_h^2 \hat{V}_{AAS}(\bar{y}_h) = \sum_{h=1}^H W_h^2 \left(\frac{1}{n_h} - \frac{1}{N_h}\right) S_h^2.$$

3.1.3. Alocação da amostra pelos estratos

A alocação da amostra pelos estratos consiste em distribuir as n unidades da amostra pelos estratos para assim garantir a precisão do procedimento amostral (Bolfarine e Bussab, 2004)

$$n = \sum_{h=1}^H n_h.$$

- Amostragem Estratificada Proporcional (AEP_P)

Consiste em distribuir proporcionalmente a amostra de tamanho n ao tamanho dos estratos (Bolfarine e Bussab, 2004),

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

Em que,

o W_h representa o peso (proporcional) do estrato h ,

N_h representa o tamanho do estrato h , e

N representa tamanho populacional, $N = \sum_{h=1}^H N_h$.

- Amostragem Estratificada Uniforme (AE_{Un})

O procedimento atribui o mesmo tamanho de amostra para cada estrato. Sendo indicado quando é de interesse apresentar estimativas para cada estrato de forma separadas (Bolfarine e Bussab, 2004),

$$n_h = \frac{n}{H} = k \text{ e } f_h = \frac{k}{N_h}.$$

Em que,

$$n = \sum_{h=1}^H n_h, \text{ e}$$

H é o total de estratos, $h = 1, \dots, H$ e

N_h representa o tamanho do estrato h .

- Alocação Ótima de Neyman (AE_{Ot})

O método considera os custos envolvidos no procedimento, onde o custo para pesquisar uma unidade amostral possa variar para cada estrato (Bolfarine e Bussab, 2004),

$$n_h = n \frac{N_h S_h / \sqrt{c_h}}{\sum_{h=1}^H N_h S_h / \sqrt{c_h}}, h = 1, \dots, H.$$

Considerando uma função de custo de forma linear,

$$C = c_0 + \sum_{h=1}^H c_h n_h,$$

sendo c_0 o custo inicial e c_h o custo por unidade observada no estrato h .

Na Amostragem estratificada para o custo fixo, pode-se calcular o tamanho da amostra, substituindo uma equação na outra resultando em

$$n = (C - c_0) \frac{\sum_{h=1}^H N_h S_h / \sqrt{c_h}}{\sum_{h=1}^H N_h S_h / \sqrt{c_h}}, h = 1, \dots, H;$$

E a variância, $V_{AE} = \sum_{h=1}^H W_h^2 \frac{S_h^2}{n_h}$.

Fixando a variância V_{AE} o tamanho da amostra é dado como

$$n = \frac{1}{V_{AE}} \left(\sum_{h=1}^H W_h S_h \sqrt{c_h} \right) \left(\sum_{h=1}^H \frac{W_h S_h}{\sqrt{c_h}} \right), h = 1, \dots, H;$$

Em que, $W_h = N_h/N$.

Considerando o custo por unidade o mesmo para todos os estratos, termos $C - c_0 = nc$, então a alocação ótima é reduzida para

$$n_h = n \frac{N_h S_h}{\sum_{h=1}^H N_h S_h}, h = 1, \dots, H.$$

E V_{AE} se reduz a $V_{ot} = \frac{1}{n} (\sum_{h=1}^H W_h S_h)^2 = \frac{\bar{S}^2}{n}$, onde $\bar{S} = \sum_{h=1}^H W_h S_h$, é o desvio padrão médio dentro de cada estrato.

3.2. AMOSTRAGEM POR CONGLOMERADOS

Os conglomerados são grupos de unidades elementares que facilitam a tarefa amostral em situações em que os sistemas de referência não são adequados e há um alto custo em realizar a atualização ou quando é cara e consome muito tempo para identificar as unidades elementares no campo (BOLFARINE; BUSSAB, 2004).

Comparando a amostragem por conglomerados (COCHRAN, 1977) com a amostragem aleatória simples de elementos, sendo ambas do mesmo tamanho, a amostragem por conglomerados tende a:

- Ter custo por elemento menor;

- Ter maior variância; e
- Maiores problemas para análises estatísticas.

Um dos problemas do uso da amostragem de conglomerados é com respeito as unidades, dentro de um mesmo conglomerado, onde estas tendem a ter valores parecidos e tornando menos eficientes estes planos. Portanto, é observado que quanto mais parecidos forem os elementos dentro do conglomerado, menos eficiente será esse método de amostragem.

Nas palavras de Bolfarie e Bussab (2004, p.162): “

[...] Para o conglomerado ser um “bom” representante do universo ele deve ser uma “micro representação” do mesmo, ou seja, ter todo tipo de participante e não do mesmo tipo. E o oposto do exigido para construção de estratos.”

3.2.1. Amostragem por Conglomerados em Um Estágio

Consiste em dividir uma população em grupos (chamados de conglomerados), feita segundo alguma(s) característica(s) conhecida(s) na população sob estudo. A divisão é feita de modo que os elementos dentro de cada conglomerados sejam diferentes entre si e que os conglomerados também o sejam entre eles. Ou seja, cada conglomerado deve ser uma representação da população como um todo. Sorteia-se um determinado número de conglomerados e, de cada um desses conglomerados sorteados, observa-se todos os seus elementos.

De forma resumida, podemos dizer que as amostras serão os elementos que compõem os conglomerados selecionados por um plano amostral qualquer (VIEIRA; SILVA, 2015).

3.2.2. Notações importantes

Neste trabalho foram adotadas as seguintes notações para amostragem por conglomerados:

O total dos conglomerados é representado como N e as unidades do conglomerado i como M_i . Todas as unidades da população são representadas como $M_0 = \sum_{i=1}^N M_i$ e todas as unidades da amostra como $m_0 = \sum_{i=1}^n M_i$.

Para o conglomerado i , o valor da variável de pesquisa para unidade j é y_{ij} , total é $y_i = \sum_{j=1}^{M_i} y_{ij}$ e a média é $\bar{Y}_i = \frac{y_i}{M_i} = \frac{\sum_{j=1}^{M_i} y_{ij}}{M_i}$.

O total da população é dado como $Y = \sum_{i=1}^N y_i$, a média por conglomerado como $\bar{Y}_C = \frac{Y}{N} = \frac{\sum_{i=1}^N y_i}{N}$ e a média por unidade como $\bar{Y} = \frac{Y}{M_0} = \frac{\sum_{i=1}^N y_i}{M_0}$.

3.2.3. Amostragem Conglomerada Simples (ACS)

Serão sorteados conglomerados, através de um processo de Amostragem Aleatória Simples com reposição – AASc – (BOLFARINE; BUSSAB, 2004, p.62) e são entrevistados todos os indivíduos de cada conglomerado. As notações para ACS serão definidas para total da população como Y , o total do conglomerado i será y_i e as unidades do conglomerado como M_i . As outras notações são dadas da seguinte forma:

Os valores amostrais são y_{ij} ; para todo $j = 1, \dots, M_i$ e $i = 1, \dots, n$. No conglomerado i da amostra o total é $y_i = \sum_{j=1}^{M_i} y_{ij}$ e a média é $\bar{Y}_i = \frac{y_i}{M_i} = \frac{\sum_{j=1}^{M_i} y_{ij}}{M_i}$.

O total amostral, $t = \sum_{i=1}^n y_i$, com média no conglomerado, $\bar{y}_C = \frac{t}{n} = \frac{\sum_{i=1}^n y_i}{n}$.

A Probabilidade de inclusão dos conglomerados é $f = \frac{n}{N}$ e de inclusão das unidades é $f = \frac{n}{N}$.

A estimação da média por unidade elementar é $\bar{y}_N = \frac{\hat{Y}}{M_0} = \frac{N}{M_0} \frac{\sum_{i=0}^n y_i}{n} = \frac{\bar{y}_C}{M}$, A variância do estimador natural (\bar{y}_N) é $V_{ACS}(\bar{y}_N) = \frac{N^2}{M_0^2} \frac{1-f}{n} S_e^2 = \frac{1}{M^2} \frac{1-f}{n} S_e^2$; O estimador da variância do estimador natural sendo, $\hat{V}_{ACS}(\bar{y}_N) = \frac{1}{M^2} \frac{1-f}{n} S_e^2$.

A estimação do total populacional é $\hat{Y} = \frac{N}{n} \sum_{i=1}^n y_i = N\bar{y}_C = \sum_{i=1}^n \sum_{j=1}^{M_i} w_{ij} \times y_{ij}$, onde os pesos individuais são dados por: $w_{ij} = \frac{N}{n}$.

A variância do estimador total populacional é $V_{ACS}(\hat{Y}) = N^2(1-f) \times \frac{S_e^2}{n}$, em que a variância entre os totais dos conglomerados é dada por $S_e^2 = \sum_{i=1}^N \frac{(y_i - \bar{y}_C)^2}{(N-1)}$. O estimador da variância: $\hat{V}_{ACS}(\hat{Y}) = N^2(1-f) \times \frac{s_e^2}{n}$; Em que, o estimador S_e^2 é dada por $s_e^2 = \sum_{i=1}^n \frac{(y_i - \bar{y}_C)^2}{(n-1)}$.

3.2.4. Amostragem por Conglomerados em Dois Estágios

Se for verificada a existência de homogeneidade interna nos conglomerados o uso de amostragem por conglomerados em um estágio não é recomendado, pois todas as suas unidades serão parecidas e terão o mesmo tipo de informação, logo ocorrerá perda de precisão. Então em uma situação hipotética em que todos os elementos são iguais, bastaria conhecer a informação de um elemento amostral para conhecer a informação de todo os elementos do conglomerado (BOLFARINE; BUSSAB, 2004, p. 197).

Uma forma de resolver tal problema sem aumentar a amostra, mas aumentando a eficiência, é subsortear os elementos dos conglomerados selecionados. Caracterizando um plano amostral em dois estágios, dado da seguinte forma:

Em população agrupada em N conglomerados: Inicialmente é sorteado no primeiro estágio n conglomerados (dado um plano amostral qualquer) e na sequência de

cada conglomerado sorteado, sorteiam-se m_i elementos de M_i elementos existentes, dado o mesmo ou outro plano amostral.

O tamanho total da amostra é dado por $m_0 = \sum_{i=1}^n m_i$.

3.2.5. Amostragem por Conglomerados em Dois Estágios com Amostragem Aleatória Simples (AC2 – AASc)

O plano amostral é indicado por AC2, onde ocorre o sorteio de n conglomerados (unidades primárias - UPA) por AASc e em seguida, também por AAS, sorteiam-se m_i elementos (unidades secundárias - USA).

Devemos observar que a notação a ser usada é a mesma da seção anterior, mas estatísticas dentro do conglomerado podem variar, quando empregamos a amostragem por conglomerados em dois estágios. Temos as seguintes notações para AC2 - AASc:

A estimação não viciada do total,

$$\hat{Y}_{HT} = \frac{N}{n} \sum_{i=1}^n M_i \bar{y}_i = \frac{N}{n} \sum_{i=1}^n \hat{y}_i$$

e variância do estimador não viciado do total:

$$V_p(\hat{Y}_{HT}) = \frac{N^2 (1-f_1)}{n (N-1)} \sum_{i=1}^N (y_i - \bar{Y}_c)^2 + \frac{N}{n} \sum_{i=1}^N M_i^2 (1 - f_{2i}) \frac{S_{2i}^2}{m_i},$$

Em que,

a fração amostral do primeiro estágio é dada por $f_1 = \frac{n}{N}$; e

a fração amostral do segundo estágio no conglomerado i é dada por

$$f_{2i} = \frac{m_i}{M_i}.$$

A variância dentro da UPA i é dada por $S_{2i}^2 = \frac{1}{M_i - 1} \sum_{j=1}^{M_i} (y_{ij} - \bar{Y}_i)^2$.

O estimador não viciado da variância é dado por

$$\hat{V}_p(\hat{Y}_{HT}) = \frac{N^2(1-f_1)}{n(N-1)} \sum_{i=1}^n (\hat{y}_i - \bar{y}_c)^2 + \frac{N}{n} \sum_{i=1}^n M_i^2 (1-f_{2i}) \frac{s_{2i}^2}{m_i}$$

Em que,

a média por conglomerado Y_c é estimada por $\bar{y}_c = \frac{1}{n} \sum_{i=1}^n \hat{y}_i$; e

a variância dentro da UPA i é estimada por: $s_{2i}^2 = \frac{1}{m_i-1} \sum_{j=1}^{m_i} (y_{ij} - \bar{y}_i)^2$.

3.3. AMOSTRAGEM COM PROBABILIDADES PROPORCIONAIS AO TAMANHO

A amostragem com Probabilidades Proporcionais ao Tamanho (PPT), tem estimação baseadas em esquemas probabilísticos mais gerais que os apresentados nas seções anteriores. Além disso, pode haver variação de tamanho nas unidades de amostragem e ao ignorar tal variação pode ocasionar em planos amostrais não eficientes.

Seja uma população com N unidades conglomeradas e a cada elemento i temos um tamanho de medida x_i . Sendo assim, temos as seguintes notações para amostragem por PPT:

Dado o parâmetro, total populacional: $Y = \sum_U Y_i$; Com a estimação linear: $\hat{Y} = \sum_s w_i y_i = \sum_U w_i y_i \delta_i$, onde w_i é o peso da unidade i e δ_i são as variáveis indicadoras de inclusão na amostra s .

Se a relação $E_p(\hat{Y}) = Y \Leftrightarrow \sum_U w_i y_i E_p(\delta_i) = \sum_U Y_i$ ocorre, temos que, o estimador \hat{Y} é não viciado para quaisquer valores y_i , sendo os pesos das unidades na amostra iguais ao inverso das respectivas probabilidades de inclusão na amostra $w_i = \pi_i^{-1}$, para todo $i \in U$. Notação utilizada:

O estimador total (não viciado),

$$\hat{Y} = \sum_s \frac{y_i}{\pi_i} = \hat{Y}_{HT}.$$

3.3.1. Estimador Horvitz-Thompson (HT)

É um estimador, não tendencioso, do total populacional que tem por objetivo trabalhar com um universo finito de amostras retiradas sem reposição com probabilidades desiguais de seleção (HORVITZ; THOMPSON, 1952, p. 663-685). Pode ser usado em qualquer plano amostral, com ou sem reposição. Comparado com planos amostrais que envolvem probabilidades iguais de seleção para as unidades amostrais o estimador possui mais eficiência ao ser empregados

Em uma população pequena onde é possível identificar seus elementos individualmente, podemos atribuir um vetor de probabilidades de seleção para esses elementos. Escolhendo uma variável que esteja relacionada com a variável de interesse, devemos calcular as probabilidades de inclusão que representam a probabilidade condicional de um dado elemento estar incluído na amostra e também para cada par de elementos a probabilidade de inclusão na amostra.

Temos então o estimador Horvitz e Thompson:

O estimador total (não viciado) é dado por $\hat{Y}_{HT} = \sum_{i=1}^n z_i$;

Onde:

A probabilidade de inclusão na amostra do *i*-ésimo elemento é π_i ;

A medida da variável de interesse do *i*-ésimo elemento é y_i .

O estimador é não tendencioso para o total Y :

Seja Z é uma variável indicadora do elemento na amostra e quando amostrado $Z_i = 1$ e quando não amostrado $Z_i = 0$. Portanto, $P(Z_i = 1) = \pi_i$, Lohr (1999). Logo obtemos, $E_p(\hat{Y}_{HT}) = E_p\left(\sum_{i=1}^n z_i \frac{y_i}{\pi_i}\right) = \sum_{i=1}^n \pi_i \frac{y_i}{\pi_i} = Y$.

A média populacional é obtida dividindo o estimador total não viciado de Horvitz e Thompson por N .

A variância associada ao estimador é dada por:

$$V_p(\hat{Y}_{HT}) = \sum_i^N \pi_i(1 - \pi_i) \left(\frac{y_i}{\pi_i}\right)^2 + \sum_{i=1}^N \sum_{j \neq i}^N (\pi_{ij} - \pi_i \pi_j) \left(\frac{y_i y_j}{\pi_i \pi_j}\right);$$

Onde:

A probabilidade de inclusão na amostra do *i*-ésimo elemento é π_i ;

A probabilidade de inclusão na amostra do *j*-ésimo elemento é π_j ;

A probabilidade de que conjuntamente os elementos *i* e *j* estejam na amostra é π_{ij} ;

A medida da variável de interesse do *i*-ésimo elemento é y_i . O estimador da variância associada ao \hat{Y}_{HT} é dada por

$$\hat{V}_p(\hat{Y}_{HT}) = \sum_i^n \frac{\pi_i(1-\pi_i)}{\pi_i} \left(\frac{y_i}{\pi_i}\right)^2 + \sum_{i=1}^n \sum_{i \neq j}^n \frac{(\pi_{ij} - \pi_i \pi_j)}{\pi_{ij}} \left(\frac{y_i y_j}{\pi_i \pi_j}\right).$$

3.4. ESTIMADOR RAZÃO

No estudo apresentado temos por objetivo estimar taxas que são razões entre totais, portanto é necessário definir (de forma resumida) o estimador razão.

Seja X a população total de x_i e Y a população total de y_i , onde X e Y são positivamente correlacionados, ou seja proporcionais, então a razão a ser estimada é dada por: $R = \frac{Y}{X}$. Sendo assim, o estimador de R é $\hat{R} = \frac{\hat{Y}}{\hat{X}}$, a razão entre os estimadores \hat{Y} e \hat{X} , que foram obtidos dado um plano amostral qualquer. (COCHRAN, 1977).

A variância aproximada da distribuição amostral linearizada de \hat{R} é obtida pelo Método Delta (CASELA; BERGER, 2001), logo a variância é dada por

$$V(\hat{R}) \approx \frac{1}{X^2} \{V(\hat{Y}) - 2RCov(\hat{Y}, \hat{X}) + R^2V(\hat{X})\}.$$

Realizando uma substituição direta de \hat{X} , \hat{R} , e as variâncias e covariâncias estimadas de \hat{X} e \hat{Y} temos o seguinte estimador da variância

$$\hat{V}(\hat{R}) = \frac{1}{\hat{X}^2} \{ \hat{V}(\hat{Y}) - 2\hat{R}Cov(\hat{Y}, \hat{X}) + \hat{R}^2 \hat{V}(\hat{X}) \}.$$

3.4.1. Estimador Razão aplicado pelo Stata

Com a finalidade de ajudar os cálculos no trabalho foi utilizado as ferramentas do software *Stata 15*, que fornece uma série de recursos computacionais. Para estimar às razões obtidas foi usado o comando *ratio*, sendo ele uma das opções do comando *svy*, que por sua vez é o prefixo do comando *survey*, comando este, usado para analisar pesquisas com plano amostral complexo (Stata, 2013, p. 65). Sendo assim, iremos mostrar como foi implementado o estimador razão no *Stata 15* no comando *ratio* (Stata, 2017, p. 5).

Seja Y_i e X_i itens de uma pesquisa com i -ésimo indivíduos na população, onde $i = 1, \dots, N$ e N é o tamanho da população. A razão da população associada para os itens de interesse é $R = \frac{Y}{X}$, onde $Y = \sum_{i=1}^N Y_i$ e $X = \sum_{i=1}^N X_i$

E se y_i e x_i são itens da pesquisa correspondentes a i -ésimos indivíduos amostrados da população, onde $i = 1, \dots, n$; Onde n é o número de observações na amostra. Então, o estimador \hat{R} para a população R é $\hat{R} = \frac{\hat{Y}}{\hat{X}}$, onde $\hat{Y} = \sum_{i=1}^n w_i y_i$ e $\hat{X} = \sum_{i=1}^n w_i x_i$ e w_i representa o peso amostral.

A variável score para o estimador razão é dada por,

$$z_i(\hat{R}) = \frac{y_i - \hat{R}x_i}{\hat{X}} = \frac{\hat{X}y_i - \hat{Y}x_i}{\hat{X}^2}.$$

Para a pós-estraficação o estimador razão é obtido pelo software *Stata 15* (Stata, 2017, p. 6) da seguinte forma:

Seja P_j o conjunto de observações amostrais que pertence ao pós-estrato j e definindo $I_{P_j}(i)$ para indicar se i -ésima observação é membro do pós-estrato j , onde $j = 1, \dots, M_p$ e M_p é o número de pós-estratos. E ainda, denotando N_j o tamanho da população para o pós-estrato j . Os elementos P_j e N_j são identificados como opções para os

comandos *poststrata ()* e *postweight ()*, respectivamente, no comando *svyset*, que será explicado mais detalhadamente na seção 4.2.

O estimador razão pós-estratificado é dado como

$$\hat{R}^P = \frac{\hat{Y}^P}{\hat{X}^P};$$

Em que,

$$\hat{Y}^P = \sum_{j=1}^{M_P} \frac{N_j}{\hat{N}_j} \hat{Y}_j = \sum_{j=1}^{M_P} \frac{N_j}{\hat{N}_j} \sum_{i=1}^n I_{P_j}(i) w_i y_i$$

e

$$\hat{X}^P = \sum_{j=1}^{M_P} \frac{N_j}{\hat{N}_j} \hat{X}_j = \sum_{j=1}^{M_P} \frac{N_j}{\hat{N}_j} \sum_{i=1}^n I_{P_j}(i) w_i x_i.$$

A variável score para o estimador razão pós-estratificado é dada por

$$z_i(\hat{R}^P) = \frac{z_i(\hat{Y}^P) - \hat{R}^P z_i(\hat{X}^P)}{(\hat{X}^P)} = \frac{\hat{X}^P z_i(\hat{Y}^P) - \hat{Y}^P z_i(\hat{X}^P)}{(\hat{X}^P)^2};$$

Em que,

$$z_i(\hat{Y}^P) = \sum_{j=1}^{M_P} I_{P_j}(i) \frac{N_j}{\hat{N}_j} \left(y_i - \frac{\hat{Y}_j}{\hat{N}_j} \right), \text{ e}$$

$$z_i(\hat{X}^P) = \sum_{j=1}^{M_P} I_{P_j}(i) \frac{N_j}{\hat{N}_j} \left(x_i - \frac{\hat{X}_j}{\hat{N}_j} \right).$$

A estimação da subpopulação pelo *Stata 15* (Stata, 2017, p. 7) foi definida da seguinte forma. Denotando S como o conjunto de observações amostrais que pertencem à subpopulação de interesse e definindo $I_S(i)$ para indicar se a i -ésima observação se enquadra na subpopulação.

O estimador razão para a subpopulação é $\hat{R}^S = \frac{\hat{Y}^S}{\hat{X}^S}$, onde

$$\hat{Y}^S = \sum_{i=1}^n I_S(i) w_i y_i, \text{ e}$$

$$\hat{X}^S = \sum_{i=1}^n I_S(i) w_i x_i.$$

A variável score é $z_j(\hat{R}^S) = I_S(i) \frac{y_i - \hat{R}^S x_i}{\hat{X}^S} = I_S(i) \frac{\hat{X}^S y_i - \hat{Y}^S x_i}{(\hat{X}^S)^2}$.

3.5. EFEITO DO PLANO AMOSTRAL (EPA)

Foi proposto por Kish (1965) uma medida para verificar o efeito do plano amostral sobre a variância de um estimador denominada Efeito do Plano Amostral (EPA). Tal medida tem por objetivo comparar a variância de um estimador qualquer com relação a outro considerado como padrão, ou seja, é uma razão entre variâncias de um estimador, comparando dois planos amostrais.

$$EPA_{Kish} = \frac{V_{Verd}(\hat{\theta})}{V_{AAS}(\hat{\theta})},$$

em que,

estimador de variância sob plano amostral verdadeiro: $V_{Verd}(\hat{\theta})$; e

estimador de variância sob AAS: $V_{AAS}(\hat{\theta})$.

Portanto, podemos saber através do EPA qual estimador é mais eficiente. Logo, o melhor estimador é aquele que apresenta a menor variabilidade.

Se $EPA_{Kish} < 1$, a variância sob AAS superestima variância verdadeira;

Se $EPA_{Kish} \cong 1$, a variância sob AAS é aproximadamente não viciada; e

Se $EPA_{Kish} > 1$, a variância sob AAS subestima variância verdadeira.

3.6. TAXAS

Uma taxa é uma relação entre duas grandezas, onde temos um coeficiente que exprime a relação existente entre uma quantidade e a frequência de um fenômeno (WENTWORTH et al, 1922). Portanto, permite expressar a presença de uma situação que não pode ser capitada, medida e nem calculada de forma direta.

No estudo aqui apresentado temos interesse nas seguintes taxas: taxa de participação na força de trabalho, nível da ocupação, taxa de ocupação, nível da desocupação, taxa de desocupação, taxa de informalidade e taxa de formalidade. Mas antes de apresentá-las devemos definir os elementos que compõem tais taxas.

Os conceitos apresentados na sequência foram estipulados pelo IBGE em suas notas metodológicas:

- Força de trabalho: são as pessoas ocupadas e as pessoas desocupadas na semana de referência (IBGE, 2014, p.17);

- Pessoas em idade de trabalhar: as pessoas de 14 anos ou mais de idade na data de referência (IBGE, 2014, p.16);

- Pessoas ocupadas: são aquelas que na semana de referência, trabalharam pelo menos uma hora completa em algum trabalho remunerado – seja em dinheiro, produtos, mercadorias ou benefícios – ou em trabalho sem remuneração direta, que ajudam à atividade econômica de membro do domicílio e pessoas que tinham trabalho remunerado e se encontravam em estado de afastamento temporário (IBGE, 2014, p.17);

- Pessoas desocupadas: são aquelas que na semana de referência estavam sem trabalho que gera rendimentos para o domicílio, que procuram emprego e estava disponível no período de referência de 30 dias. Pessoas sem trabalho na semana de referência, mas que já haviam conseguido trabalho e começariam trabalhar após a semana de referência, também são consideradas desocupadas (IBGE, 2014, p.17);

- Trabalho formal: é o trabalho realizado com carteira assinada, bem como o trabalho domésticos também com carteira assinada, o trabalho de militares e funcionários públicos estatutários, e ainda, empregadores e trabalhadores por conta própria que contribuem para a previdência social (IBGE, 2016, p.132);

- Trabalho informal: é o trabalho realizado sem carteira assinada, incluindo os trabalhadores domésticos, empregadores e trabalhadores por conta própria que não contribuem para a previdência social, trabalhadores não remunerados, bem como os trabalhadores na produção para o próprio consumo e na construção para o próprio uso (IBGE, 2016, p.132).

Definidos os conceitos vamos as taxas do estudo.

- Taxa de participação na força de trabalho: É o percentual de pessoas na força de trabalho, em relação às pessoas em idade de trabalhar, na semana de referência,

$$\frac{\textit{Total de pessoas na força de trabalho}}{\textit{Total de pessoas em idade de trabalhar}} \times 100\%.$$

- Nível da ocupação: É o percentual de pessoas ocupadas em relação às pessoas em idade de trabalhar na semana de referência,

$$\frac{\textit{Total de pessoas ocupadas}}{\textit{Total de pessoas em idade de trabalhar}} \times 100\%.$$

- Taxa de ocupação: É o percentual de pessoas ocupadas em relação às pessoas na força de trabalho na semana de referência,

$$\frac{\textit{Total de pessoas ocupadas}}{\textit{Total de pessoas na força de trabalho}} \times 100\%.$$

- Nível da desocupação: É o percentual de pessoas desocupadas, em relação às pessoas em idade de trabalhar na semana de referência,

$$\frac{\textit{Total de pessoas desocupadas}}{\textit{Total de pessoas em idade de trabalhar}} \times 100\%.$$

- Taxa de desocupação: É o percentual de pessoas desocupadas, em relação às pessoas na força de trabalho na semana de referência,

$$\frac{\textit{Total de pessoas desocupadas}}{\textit{Total de pessoas na força de trabalho}} \times 100\%.$$

- Taxa de informalidade: Pode ser definida de mais de uma forma, e na monografia aqui apresentada, usaremos a forma mais utilizada entre os analistas de mercado do trabalho, onde um trabalhador dado como por conta-própria para ser considerado informal, é necessário que ele não tenha contribuído para a previdência social (IPEA, 2017, p.17). Logo, a taxa de informalidade é dada pelo percentual de trabalhadores informais, em relação a soma dos trabalhadores formais e informais,

$$\frac{\textit{Total de trabalhadores informais}}{\textit{Total de trabalhadores formais} + \textit{Total de trabalhadores informais}} \times 100\%.$$

- Taxa de formalidade: é o percentual de trabalhadores formais, em relação a soma dos trabalhadores formais e informais,

$$\frac{\textit{Total de trabalhadores formais}}{\textit{Total de trabalhadores formais} + \textit{Total de trabalhadores informais}} \times 100\%.$$

4 – EPA NA PNAD CONTÍNUA

4.1. BANCO DE DADOS

O banco de dados da PNAD Contínua está disponível no site do IBGE e para a análise desta monografia o banco de dados foi obtido através do laboratório ECONS: <http://www.ufjf.br/econs/bancosdedados/documentacao-dos-dados/> da Faculdade de Economia da Universidade Federal de Juiz de Fora (UFJF).

A PNAD Contínua possui um banco de dados composto por entrevistas presenciais e com isso é extenso e complexo abrangendo vários aspectos sociais e econômicos dos participantes da pesquisa. Foram analisadas informações que se referem a taxas sobre o tema trabalho para os trimestres do período de 2012 a 2017.

As informações para o tema citado são captadas em dois questionários: um reduzido, onde as informações são restritas ao trabalho, que gera rendimentos para o domicílio (cuja produção é voltada para o mercado) e um ampliado, mais abrangente, aplicado na primeira entrevista de cada domicílio, que além das variáveis contidas no reduzido, inclui informações sobre outras formas de trabalho, quando a produção não é voltada para o mercado.

4.2. VARIÁVEIS

Para análise do desenho amostral da PNADC foram usadas as variáveis indicadas pelo IBGE do Dicionário de Variáveis da pesquisa (disponível no site indicado na seção Referências), ver Quadro I, e a estruturação indicada por Pessoa et al. (1997), já inserindo as variáveis no contexto do plano amostral da PNADC.

O comando *svyset* para reproduzir o plano amostral da PNAD Contínua exige as seguintes variáveis (StataCorp, 2013, p. 165):

```
svyset [psu] [weight], strata(varname) single(varname) posts(varname)
```

- *psu*: a variável que representa a unidade primária de amostragem;
- *weight*: a variável que representa o peso a ser atribuído a cada observação;
- *strata*: a variável que represente o estrato;
- *single*: especifica como manipular estratos com uma única unidade de amostragem. O Stata nos dá as opções:

missing (método padrão do programa), resulta em valores ausentes para os erros-padrão;

certainty, faz com que estratos com unidades de amostragem únicas sejam tratados como unidades de certeza. Unidades de certeza não contribuem com nada para o erro padrão;

scaled, resulta em uma versão escalada da *single(certainty)*. O fator de escala resulta da utilização da média dos desvios dos estratos com várias unidades de amostragem para cada estrato com uma unidade de amostragem; e

centered, especifica que os estratos com uma unidade de amostragem estão centrados na média geral em vez da média do estrato.

Para repetir o plano amostral da PNAD Contínua é escolhido *centered*, pois em pesquisas por amostragem a estimação de variância envolve cálculos de desvios em unidades primárias de amostragem. Se apenas uma informação é amostrada em um estrato particular a variância não pode ser computada no *Stata*, o *svyset* tem como padrão considerar como *missing* esta opção, mas podemos modificada pela opção *single()*. Como estamos interessados em implementar o plano amostral da PNAD-C usaremos a opção *single(centered)*, que estima as variâncias para as observações únicas por estrato como centradas na média da amostra.

- *posts*: a variável que indica os pós-estratos; e

- *postw*: a variável que indica o peso da população do pós-estrato.

Com isso, retiramos do dicionário de variáveis da PNAD Contínua a variável UPA, que representa a unidade primária da amostra e é inserida no comando *psu()*; A variável ESTRATO, que representa os estratos de seleção amostral é inserida no comando *strata()*; A variável V1027 representa o peso é inserida no comando *weight()*. Na sequência para corrigir os pesos para não resposta a PNAD Contínua disponibiliza a variável V1029, que será inserida no comando *postw()* e representa a projeção da população para cada POSEST. Por sua vez, POSEST será inserida no comando *posts()*, representa os estratos para correções de não resposta em que são feitas as projeções da população.

Quadro 1 – Variáveis usada do dicionário de variáveis da PNAD Contínua 2017

Código da variável	Quesito	Categorias	
	descrição	Tipo	Descrição
UPA	Unidade Primária de Amostragem (UPA)		UF (2) + Número Sequencial (6) + DV (1)
ESTRATO			As 2 primeiras posições representam o código da Unidade da Federação
V1027 ¹	Peso do domicílio e das pessoas	6 dígitos e 8 casas decimais	Peso trimestral com correção de não entrevista sem pós estratificação pela projeção de população
V1029	Projeção da população		Projeção da população do trimestre (referência: mês do meio)
POSEST	Domínios de projeção		As 2 primeiras posições representam o código da Unidade da Federação e a última, o tipo de área. UF(2) + V1023(1)
V2009	Idade do morador na data de referência	0 a 130	Idade (em anos)
VD4001	Condição em relação à força de trabalho na semana de referência para pessoas de 14 anos ou mais de idade	1	Pessoas na força de trabalho
		2	Pessoas fora da força de trabalho
			Não aplicável
VD4002	Condição de ocupação na semana de referência para pessoas de 14 anos ou mais de idade	1	Pessoas ocupadas
		2	Pessoas desocupadas
			Não aplicável

Fonte: IBGE – Dicionário de Variáveis da PNAD Contínua 2012-2017.

Também foram criadas variáveis para obter as taxas e seus respectivos intervalos de confiança com 95% de confiabilidade e erro, tanto para o desenho amostral complexo da PNAD Contínua, quanto para o desenho amostral sob AAS, além de uma variável com o objetivo de obter o efeito do plano amostral (EPA), ver Quadro 2.

¹ A variável V1027 foi estabelecida sem as correções de não respostas, além ser o inverso da probabilidade de seleção de cada elemento da amostra, logo ao invés de corrigimos os pesos de não resposta, corrigimos as estruturas de estimação para incorporar a possibilidade de não respostas pelos pacotes fornecidos pelo software *Stata*[®] 15 (StataCorp, 2017) com a função *postw()* (West et al., 2009).

Quadro 2 - Variáveis criadas para o estudo.

Variável	Descrição
Ano	Indica o ano de 2012 a 2017.
Trimestre	Indica 1º ao 4º trimestre.
LInf1	Limite inferior da taxa ou nível obtido sob o plano amotral da PNAD Contínua.
LSup1	Limite superior da taxa ou nível obtido sob o plano amotral da PNAD Contínua.
Erro1	Erro da taxa ou nível obtido sob o plano amotral da PNAD Contínua.
LInf2	Limite inferior da taxa ou nível obtido sob o plano amotral AAS.
LSup2	Limite superior da taxa ou nível obtido sob o plano amotral AAS.
Erro2	Erro da taxa ou nível obtido sob o plano amotral AAS.
EPA	Efeito do Plano Amotral (EPA) obtido de cada taxa analisada.

Fonte: Elaboração própria para análise da PNAD Contínua 2012-2017.

Sendo assim, podemos acompanhar os valores do EPAs trimestralmente de 2012-2017 para o Brasil. Lembrando que, não temos por objetivo nesta monografia discutir os motivos econômicos que levaram a tais taxas e sim manter o foco em analisar as questões metodológicas do plano amostral que resulta nos indicadores obtidos através da PNADC para o Brasil.

- Definições dos intervalos de confiança e erro

Para as estimações dos intervalos de confiança e erros-padrão foi utilizado ferramentas do software *Stata*[®] 15, através do comando *ratio*, que para realizar o cálculo utiliza o teste para proporção de uma única amostra, possui o nível de significância com default no valor 95% de confiabilidade (StataCorp, 2017, p. 2075-2077). O teste usado pelo programa é definido da seguinte forma.

Pelo Teorema do Limite Central (Ross, 2010, p. 463) \bar{X} possui aproximadamente uma distribuição normal, ou seja, $\bar{X} \sim N\left(p, \frac{p(1-p)}{n}\right)$.

E como \bar{X} é um estimador de máximo verossimilhança para a proporção populacional p e para um n suficientemente grande consideramos a distribuição amostral $\hat{p} = \bar{X}$ como aproximadamente normal, ou seja, $\hat{p} \sim N\left(p, \frac{p(1-p)}{n}\right)$. Onde n é o número de observações, \hat{p} a proporção observada e $\hat{q} = 1 - \hat{p}$.

$$\text{Logo, } Z = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} \sim N(0,1).$$

Estabelecendo a hipótese bicaudal

$$\begin{cases} H_0: p = p_0 \\ H_1: p \neq p_0. \end{cases}$$

Temos assim o teste bicaudal de uma proporção populacional usando uma estatística de teste assintoticamente distribuída normalmente dada por,

$$Z_{obs} = \frac{\hat{p} - p_0}{s_0}.$$

Onde p_0 é a proporção hipotética, $q_0 = 1 - p_0$ e o $s_0 = \sqrt{\frac{p_0 q_0}{n}}$ é o erro padrão de \hat{p} sob a hipótese nula de $p = p_0$ para um nível de significância α .

O intervalo de confiança é dado por

$$IC(p, 1 - \alpha) = \left(\hat{p} - Z_{\alpha/2} \sqrt{\hat{p}, \frac{\hat{p}(1-\hat{p})}{n}}; \hat{p} + Z_{\alpha/2} \sqrt{\hat{p}, \frac{\hat{p}(1-\hat{p})}{n}} \right).$$

4.3. RESULTADOS

Usando as informações fornecidos pelo banco de dados da PNAD Contínua de 2012 a 2017 foram produzidas estimativas para o Brasil no que diz respeito a taxa de participação da força de trabalho, o nível de ocupação, a taxa de ocupação, o nível de desocupação, a taxa de desocupação, a taxa de formalidade e taxa de informalidade.

Os resultados foram agrupados em Tabelas que informam as estimativas médias, intervalos de confiança e erro padrão. Foram exibidos de forma trimestral para cada uma das taxas estudadas, tanto para o desenho amostral da PNADC quanto para o desenho amostral mais ingênuo sob AAS. Por fim, há a inserção de uma coluna que representa as estimativas para o EPA de cada um dos trimestres de 2012-2017 e para cada uma das taxas apresentadas.

Lembrando que, valores afastados de um, o EPA indica que as estimativas são viciadas e incorretas, pois foi ignorado o plano amostral na estimação da variância. E ainda, valores grandes indicam que houve a subestimação da variância verdadeira do estimador, dado que o estimador da variância foi obtido ignorando o plano amostral complexo (SILVA et al, 2002, p. 19).

Na Tabela 1 estão os resultados obtidos através da base de dados da PNADC para a taxa de participação da força de trabalho de 2012 a 2017 e exibem oscilações relativamente pequenas em torno da média geral de 61,35%, com crescimento entre o primeiro trimestre de 2015 e o segundo trimestre de 2016 (de 61,02% para 61,58%). No terceiro trimestre de 2016, houve uma redução da taxa de participação para 61,18% e apresentou uma elevação sucessiva com uma leve queda no último trimestre de 2017 com 61,77%.

Ao considerar o plano amostral sob AAS verificamos que as médias a cada trimestre ao longo do período referido são inferiores ao plano amostral da PNADC para o mesmo período, assim como os erros-padrão respectivos, ocasionado em limites superiores e inferiores mais curtos, portanto temos um intervalo de confiança menor sob AAS. Além disso, temos um EPA médio de 2,98, indicando ser um erro ignorar o plano amostral e assim subestimando a média, pois há indícios no processo de estimação de que os efeitos de conglomeração são mais fortes do que os efeitos da estratificação.

Tabela 1: Resultados para a Taxa Participação da Força de Trabalho (Brasil – PNADC 2012-2017)

Ano	Trimestre	Estimativa Plano amostral PNADC				Estimativa Plano amostral sob AAS				EPA estimado
		Média (%)	Limite Inferior (%)	Limite Superior (%)	Erro padrão	Média (%)	Limite Inferior (%)	Limite Superior (%)	Erro padrão	
2012	1º	61,16	60,92	61,40	0,0012	60,06	59,91	60,20	0,0007	2,71
2012	2º	61,70	61,47	61,94	0,0012	60,63	60,49	60,78	0,0007	2,58
2012	3º	61,54	61,30	61,77	0,0012	60,42	60,28	60,57	0,0007	2,66
2012	4º	61,29	61,05	61,52	0,0012	60,18	60,03	60,32	0,0007	2,62
2013	1º	61,18	60,95	61,42	0,0012	60,10	59,96	60,25	0,0007	2,63
2013	2º	61,49	61,26	61,73	0,0012	60,18	60,04	60,33	0,0007	2,63
2013	3º	61,35	61,12	61,59	0,0012	60,05	59,91	60,20	0,0007	2,64
2013	4º	61,05	60,82	61,29	0,0012	59,81	59,66	59,95	0,0007	2,70
2014	1º	61,14	60,89	61,39	0,0013	59,72	59,58	59,87	0,0007	2,99
2014	2º	61,10	60,86	61,34	0,0012	59,65	59,50	59,79	0,0007	2,88
2014	3º	60,93	60,69	61,16	0,0012	59,54	59,40	59,68	0,0007	2,69
2014	4º	60,88	60,63	61,13	0,0013	59,37	59,22	59,51	0,0007	2,98
2015	1º	61,02	60,77	61,27	0,0013	59,49	59,35	59,63	0,0007	3,04
2015	2º	61,28	61,03	61,53	0,0013	59,63	59,49	59,77	0,0007	3,14
2015	3º	61,44	61,19	61,69	0,0013	59,70	59,56	59,84	0,0007	3,06
2015	4º	61,42	61,16	61,69	0,0014	59,45	59,30	59,59	0,0007	3,39
2016	1º	61,44	61,18	61,71	0,0014	59,42	59,27	59,56	0,0007	3,45
2016	2º	61,58	61,31	61,85	0,0014	59,34	59,20	59,49	0,0007	3,55
2016	3º	61,18	60,91	61,44	0,0013	58,67	58,53	58,81	0,0007	3,42
2016	4º	61,39	61,12	61,65	0,0013	58,79	58,65	58,94	0,0007	3,33
2017	1º	61,55	61,29	61,81	0,0013	58,82	58,68	58,96	0,0007	3,35
2017	2º	61,69	61,45	61,93	0,0012	59,06	58,92	59,20	0,0007	2,91
2017	3º	61,79	61,54	62,04	0,0013	59,13	58,98	59,27	0,0007	3,07
2017	4º	61,77	61,52	62,02	0,0013	59,01	58,86	59,15	0,0007	3,03

Fonte: Elaboração própria com base na PNAD Contínua 2012-2017

A seguir, na Tabela 2, temos os resultados para o nível de ocupação do Brasil de 2012 a 2017, onde a cada trimestre há um crescimento do nível de ocupação e a partir do terceiro trimestre de 2014 essa sazonalidade se inverte, mantendo essa tendência até 2016 é interessante também observar um dado trimestre ano a ano e notar a tendência de queda.

Podemos perceber que o nível médio sob AAS apresentou valores inferiores àqueles produzidos sob o desenho amostral da PNAD Contínua, tal como o intervalo de confiança menor e o EPA médio que obtemos de 2,94, indicando que os efeitos de conglomeramento são mais fortes do que os efeitos da estratificação, portanto há subestimação da nível médio, sendo assim, inadequada ignorar o plano amostral.

Tabela 2: Resultados para o Nível de Ocupação (Brasil – PNADC 2012-2017)

Ano	Trimestre	Estimativa Plano amostral PNADC				Estimativa Plano amostral sob AAS				EPA estimado
		Média (%)	Limite Inferior (%)	Limite Superior (%)	Erro padrão	Média (%)	Limite Inferior (%)	Limite Superior (%)	Erro padrão	
2012	1º	56,30	56,06	56,54	0,0012	55,46	55,32	55,61	0,0008	2,62
2012	2º	57,06	56,83	57,29	0,0012	56,23	56,08	56,38	0,0007	2,45
2012	3º	57,18	56,95	57,42	0,0012	56,31	56,16	56,46	0,0007	2,57
2012	4º	57,08	56,85	57,32	0,0012	56,20	56,05	56,35	0,0008	2,57
2013	1º	56,30	56,07	56,53	0,0012	55,44	55,30	55,59	0,0007	2,48
2013	2º	56,92	56,69	57,16	0,0012	55,88	55,73	56,02	0,0007	2,56
2013	3º	57,10	56,86	57,33	0,0012	56,10	55,96	56,25	0,0007	2,60
2013	4º	57,28	57,04	57,51	0,0012	56,29	56,15	56,44	0,0007	2,61
2014	1º	56,75	56,51	57,00	0,0013	55,71	55,56	55,85	0,0007	2,92
2014	2º	56,92	56,68	57,16	0,0012	55,79	55,65	55,94	0,0007	2,76
2014	3º	56,80	56,56	57,04	0,0012	55,77	55,63	55,91	0,0007	2,71
2014	4º	56,93	56,68	57,17	0,0012	55,80	55,65	55,94	0,0007	2,87
2015	1º	56,18	55,93	56,42	0,0012	55,05	54,91	55,20	0,0007	2,86
2015	2º	56,19	55,95	56,43	0,0012	55,06	54,92	55,21	0,0007	2,85
2015	3º	55,98	55,73	56,23	0,0013	54,85	54,71	54,99	0,0007	2,91
2015	4º	55,92	55,66	56,18	0,0013	54,60	54,46	54,75	0,0007	3,13
2016	1º	54,74	54,48	55,01	0,0014	53,46	53,31	53,60	0,0007	3,41
2016	2º	54,61	54,33	54,89	0,0014	53,07	52,93	53,22	0,0007	3,67
2016	3º	53,95	53,68	54,23	0,0014	52,20	52,06	52,35	0,0007	3,68
2016	4º	54,00	53,73	54,27	0,0014	52,11	51,96	52,25	0,0007	3,48
2017	1º	53,09	52,82	53,36	0,0014	51,20	51,06	51,35	0,0007	3,46
2017	2º	53,67	53,41	53,92	0,0013	51,81	51,67	51,96	0,0007	3,08
2017	3º	54,11	53,85	54,37	0,0013	52,22	52,07	52,36	0,0007	3,14
2017	4º	54,48	54,23	54,74	0,0013	52,47	52,32	52,61	0,0007	3,13

Fonte: Elaboração própria com base na PNAD Contínua 2012-2017

Temos agora a Tabela 3, que se refere a taxa de ocupação do Brasil, os resultados encontrados para as taxas médias para os trimestres do período de 2012 a 2017. Vemos que para o período de 2012 a 2015 as taxas apresentaram valores acima dos 91%, tendo o último trimestre de 2016 e os três primeiros trimestres de 2017 como os mais inferiores da série, ficando abaixo de 87,97% de ocupação. Comparando um dado trimestre ano a ano é perceptível a queda na taxa de ocupação do Brasil de 2012 a 2017.

No que diz respeito ao plano amostral, apresenta valores ligeiramente superiores sob AAS em relação ao desenho amostral da PNAD Contínua e o EPA médio é elevado com valor de 3,13 – esse valor é impactado principalmente pelos EPAs do terceiro trimestre de 2015 em diante – e assim mostrou ser inadequado ignorar o plano amostral e ainda apresenta fortes indícios de que efeitos de conglomeramento são mais fortes do que os efeitos da estratificação.

Tabela 3: Resultados para a Taxa de Ocupação (Brasil – PNADC 2012-2017)

Ano	Trimestre	Estimativa Plano amostral PNADC				Estimativa Plano amostral sob AAS				EPA estimado
		Média (%)	Limite Inferior (%)	Limite Superior (%)	Erro padrão	Média (%)	Limite Inferior (%)	Limite Superior (%)	Erro padrão	
2012	1º	92,05	91,89	92,21	0,0008	92,35	92,25	92,45	0,0005	2,47
2012	2º	92,48	92,32	92,63	0,0008	92,74	92,64	92,84	0,0005	2,64
2012	3º	92,93	92,77	93,08	0,0008	93,20	93,10	93,29	0,0005	2,69
2012	4º	93,14	92,99	93,29	0,0008	93,39	93,30	93,49	0,0005	2,56
2013	1º	92,02	91,86	92,18	0,0008	92,25	92,15	92,35	0,0005	2,49
2013	2º	92,57	92,41	92,72	0,0008	92,84	92,74	92,94	0,0005	2,52
2013	3º	93,06	92,91	93,22	0,0008	93,42	93,33	93,52	0,0005	2,67
2013	4º	93,82	93,68	93,96	0,0007	94,12	94,03	94,21	0,0005	2,64
2014	1º	92,83	92,66	93,00	0,0009	93,28	93,18	93,37	0,0005	3,36
2014	2º	93,15	93,00	93,30	0,0008	93,54	93,44	93,63	0,0005	2,68
2014	3º	93,23	93,07	93,38	0,0008	93,67	93,58	93,76	0,0005	2,98
2014	4º	93,50	93,35	93,66	0,0008	93,99	93,90	94,08	0,0005	2,90
2015	1º	92,06	91,90	92,22	0,0008	92,54	92,44	92,64	0,0005	2,67
2015	2º	91,69	91,52	91,87	0,0009	92,34	92,24	92,44	0,0005	3,06
2015	3º	91,12	90,94	91,30	0,0009	91,87	91,77	91,98	0,0005	3,07
2015	4º	91,04	90,85	91,24	0,0010	91,85	91,75	91,96	0,0005	3,37
2016	1º	89,10	88,84	89,36	0,0013	89,97	89,86	90,08	0,0006	5,16
2016	2º	88,68	88,46	88,91	0,0012	89,43	89,31	89,55	0,0006	3,89
2016	3º	88,20	87,97	88,43	0,0012	88,98	88,86	89,10	0,0006	3,78
2016	4º	87,97	87,75	88,19	0,0011	88,63	88,51	88,75	0,0006	3,47
2017	1º	86,25	86,00	86,51	0,0013	87,05	86,93	87,18	0,0006	3,95
2017	2º	87,00	86,77	87,22	0,0011	87,73	87,61	87,86	0,0006	3,26
2017	3º	87,57	87,35	87,79	0,0011	88,32	88,20	88,44	0,0006	3,31
2017	4º	88,21	87,99	88,43	0,0011	88,91	88,80	89,03	0,0006	3,44

Fonte: Elaboração própria com base na PNAD Contínua 2012-2017

Na sequência é apresentado a Tabela 4 que se refere ao nível de desocupação do Brasil e os resultados foram obtidos para o período de 2012 a 2017. Do período de 2012 a 2014 o nível médio apresentou pequenas oscilações em torno da média geral de 4,35%, mas a partir de 2015 apresentou elevação substancial com atenção para o primeiro trimestre de 2017 com um nível de desocupação 8,46.

O plano amostral sob AAS apresentou nível médio com valores inferiores ao plano amostral da PNAD Contínua e é possível verificar que os erros no plano amostral mais ingênuo ocasionaram um intervalo de confiança mais estreito do que o plano amostral da PNAD-C. O EPA médio também é elevado com um valor estimado de 3,13, portanto é inadequado ignorar o plano amostral, além de apresentar indícios de que há efeitos de conglomeramento e estes são mais fortes do que os efeitos da estratificação.

Tabela 4: Resultados para o Nível de Desocupação (Brasil – PNADC 2012-2017)

Ano	Trimestre	Estimativa Plano amostral PNADC				Estimativa Plano amostral sob AAS				EPA estimado
		Média (%)	Limite Inferior (%)	Limite Superior (%)	Erro padrão	Média (%)	Limite Inferior (%)	Limite Superior (%)	Erro padrão	
2012	1º	4,86	4,76	4,96	0,0005	4,59	4,53	4,65	0,0003	2,59
2012	2º	4,64	4,54	4,74	0,0005	4,40	4,34	4,46	0,0003	2,80
2012	3º	4,35	4,25	4,45	0,0005	4,11	4,05	4,17	0,0003	2,84
2012	4º	4,21	4,11	4,30	0,0005	3,98	3,92	4,04	0,0003	2,68
2013	1º	4,88	4,78	4,98	0,0005	4,66	4,60	4,72	0,0003	2,65
2013	2º	4,57	4,47	4,67	0,0005	4,31	4,25	4,37	0,0003	2,66
2013	3º	4,26	4,16	4,35	0,0005	3,95	3,89	4,01	0,0003	2,81
2013	4º	3,77	3,68	3,86	0,0005	3,51	3,46	3,57	0,0003	2,79
2014	1º	4,38	4,28	4,49	0,0006	4,02	3,96	4,07	0,0003	3,58
2014	2º	4,18	4,09	4,28	0,0005	3,86	3,80	3,91	0,0003	2,87
2014	3º	4,13	4,03	4,23	0,0005	3,77	3,71	3,82	0,0003	3,13
2014	4º	3,95	3,86	4,05	0,0005	3,57	3,52	3,62	0,0003	3,11
2015	1º	4,84	4,74	4,94	0,0005	4,44	4,38	4,50	0,0003	2,89
2015	2º	5,09	4,98	5,20	0,0006	4,57	4,51	4,63	0,0003	3,38
2015	3º	5,46	5,34	5,57	0,0006	4,85	4,79	4,91	0,0003	3,34
2015	4º	5,50	5,38	5,62	0,0006	4,84	4,78	4,91	0,0003	3,75
2016	1º	6,70	6,53	6,86	0,0008	5,96	5,89	6,03	0,0004	5,69
2016	2º	6,97	6,82	7,11	0,0007	6,27	6,20	6,34	0,0004	4,18
2016	3º	7,22	7,08	7,36	0,0007	6,47	6,39	6,54	0,0004	4,01
2016	4º	7,38	7,24	7,52	0,0007	6,68	6,61	6,76	0,0004	3,74
2017	1º	8,46	8,30	8,62	0,0008	7,62	7,54	7,69	0,0004	4,35
2017	2º	8,02	7,88	8,16	0,0007	7,24	7,17	7,32	0,0004	3,52
2017	3º	7,68	7,54	7,82	0,0007	6,91	6,83	6,98	0,0004	3,62
2017	4º	7,28	7,14	7,42	0,0007	6,54	6,47	6,61	0,0004	3,77

Fonte: Elaboração própria com base na PNAD Contínua 2012-2017

A Tabela 5 se refere a taxa de desocupação do Brasil (talvez a taxa de maior interesse, visto o impacto que gera na sociedade) e os resultados são referentes ao período de 2012 a 2017. Do período de 2012 a 2014 a taxa de desocupação apresentou pequenas oscilações em torno da média geral de 4,35%, mas a partir de 2015 apresentou elevação substancial, dada a atenção para o primeiro trimestre de 2017, que possui um percentual de 13,75% de desempregados.

Comparando um dado trimestre de 2012 com o mesmo trimestre dos anos subsequentes é notório a elevação da taxa média de desemprego, por exemplo, no primeiro trimestre de 2012 a taxa média de desemprego estava 7,95%, já no primeiro trimestre de 2017 a taxa média estava em 13,75%.

Observado o mesmo período para plano amostral sob AAS e para o plano amostral da PNADC, verificamos que as taxas médias de desocupação a cada trimestre ao longo dos anos são inferiores no plano amostral sob AAS. Os erros também são inferiores no plano amostral sob AAS, levando a produzir intervalos de confiança mais curtos.

O EPA médio observado possui um valor de 3,13, indicando ser inadequado ignorar o plano amostral e assim havendo uma subestimação da média estimada, portanto foi verificado indícios no processo de estimação de que efeitos de conglomeração são mais fortes do que os efeitos da estratificação.

Tabela 5: Resultados para a Taxa de Desocupação (Brasil – PNADC 2012-2017)

Ano	Trimestre	Estimativa Plano amostral PNADC				Estimativa Plano amostral sob AAS				EPA estimado
		Média (%)	Limite Inferior (%)	Limite Superior (%)	Erro padrão	Média (%)	Limite Inferior (%)	Limite Superior (%)	Erro padrão	
2012	1º	7,95	7,79	8,11	0,0008	7,65	7,55	7,75	0,0005	2,47
2012	2º	7,52	7,37	7,68	0,0008	7,26	7,16	7,36	0,0005	2,64
2012	3º	7,07	6,92	7,23	0,0008	6,80	6,71	6,90	0,0005	2,69
2012	4º	6,86	6,71	7,01	0,0008	6,61	6,51	6,70	0,0005	2,56
2013	1º	7,98	7,82	8,14	0,0008	7,75	7,65	7,85	0,0005	2,49
2013	2º	7,43	7,28	7,59	0,0008	7,16	7,06	7,26	0,0005	2,52
2013	3º	6,94	6,78	7,09	0,0008	6,58	6,48	6,67	0,0005	2,67
2013	4º	6,18	6,04	6,32	0,0007	5,88	5,79	5,97	0,0005	2,64
2014	1º	7,17	7,00	7,34	0,0009	6,72	6,63	6,82	0,0005	3,36
2014	2º	6,85	6,70	7,00	0,0008	6,46	6,37	6,56	0,0005	2,68
2014	3º	6,77	6,62	6,93	0,0008	6,33	6,24	6,42	0,0005	2,98
2014	4º	6,50	6,34	6,65	0,0008	6,01	5,92	6,10	0,0005	2,90
2015	1º	7,94	7,78	8,10	0,0008	7,46	7,36	7,56	0,0005	2,67
2015	2º	8,31	8,13	8,48	0,0009	7,66	7,56	7,76	0,0005	3,06
2015	3º	8,88	8,70	9,06	0,0009	8,13	8,02	8,23	0,0005	3,07
2015	4º	8,96	8,76	9,15	0,0010	8,15	8,04	8,25	0,0005	3,37
2016	1º	10,90	10,64	11,16	0,0013	10,03	9,92	10,14	0,0006	5,16
2016	2º	11,32	11,09	11,54	0,0012	10,57	10,45	10,69	0,0006	3,89
2016	3º	11,80	11,57	12,03	0,0012	11,02	10,90	11,14	0,0006	3,78
2016	4º	12,03	11,81	12,25	0,0011	11,37	11,25	11,49	0,0006	3,47
2017	1º	13,75	13,49	14,00	0,0013	12,95	12,82	13,07	0,0006	3,95
2017	2º	13,00	12,78	13,23	0,0011	12,27	12,14	12,39	0,0006	3,26
2017	3º	12,43	12,21	12,65	0,0011	11,68	11,56	11,80	0,0006	3,31
2017	4º	11,79	11,57	12,01	0,0011	11,09	10,97	11,20	0,0006	3,44

Fonte: Elaboração própria com base na PNAD Contínua 2012-2017

Nas Tabelas 6 e 7 são apresentados os valores com as estimativas para as taxas de trabalho informal e formal respectivamente, que somados representam a população brasileira caracterizado como ocupada.

De início, vamos analisar de forma breve a taxa de informalidade do Brasil. Os valores informados pela Tabela 6 mostraram para o período de 2012 a 2016 taxas médias de informalidade com um declínio constante, ficando mais evidente, quando comparamos um trimestre de um determinado ano com o mesmo trimestre do ano subsequente. Para o ano 2017 vemos uma elevação significativa do trabalho informal, novamente se houver a comparação trimestre a trimestre com o ano de 2016 temos uma percepção mais visível, mas independente do trimestre observado para qualquer um dos anos apresentados foi visível a grande parcela de informais do Brasil. Essa taxa possui dentre suas definições não ser contribuinte da previdência social, o que demonstra o importante impacto que ocorre na arrecadação previdenciária (MENEZES; DEDECCA, 2012, p. 39).

Os efeitos do plano amostral obtidos demonstraram que ignorar o plano amostral é inadequado, assim havendo uma subestimação da média estimada, dado o EPA estimado médio de 3,07. Com isso, temos indícios que o processo de estimação possui efeitos de conglomeração mais fortes que os efeitos da estratificação.

Por fim temos a Tabela 7, que mostra as estimativas para a taxa de trabalho formal de 2012 a 2017, fazendo um contraponto com a Tabela 6, nessa Tabela as taxas crescentes de trabalhadores no mercado formal seguem até 2016. A partir de 2017 observamos uma queda do emprego formal.

Assim como foi verificando em outras taxas, levando em conta o mesmo período, também observamos taxas médias de emprego formal menores no plano amostral sob AAS e os erros-padrão sempre ocasionarem em intervalos de confiança menores do que o plano amostral da PNADC. O EPA médio apresentou um valor 3,08, mostrando ser inadequado ignorar o plano amostral, subestimando a média e tendo indícios de que o processo de estimação possui efeitos de conglomeração mais fortes do que os efeitos da estratificação.

Tabela 6: Resultados para a Taxa de Informalidade (Brasil – PNADC 2012-2017)

Ano	Trimestre	Estimativa Plano amostral PNADC				Estimativa Plano amostral sob AAS				EPA estimado
		Média (%)	Limite Inferior (%)	Limite Superior (%)	Erro padrão	Média (%)	Limite Inferior (%)	Limite Superior (%)	Erro padrão	
2012	1º	41,70	41,37	42,04	0,0017	46,89	46,69	47,09	0,0010	2,86
2012	2º	41,75	41,40	42,09	0,0017	46,91	46,72	47,11	0,0010	3,02
2012	3º	41,37	41,04	41,70	0,0017	46,73	46,53	46,93	0,0010	2,84
2012	4º	41,18	40,84	41,52	0,0017	46,75	46,55	46,95	0,0010	2,93
2013	1º	40,77	40,43	41,11	0,0017	46,15	45,96	46,35	0,0010	2,94
2013	2º	40,67	40,34	41,01	0,0017	46,09	45,90	46,29	0,0010	2,91
2013	3º	40,40	40,07	40,73	0,0017	45,97	45,78	46,17	0,0010	2,80
2013	4º	40,28	39,95	40,61	0,0017	45,95	45,75	46,14	0,0010	2,86
2014	1º	39,21	38,88	39,55	0,0017	45,24	45,04	45,43	0,0010	2,91
2014	2º	38,91	38,58	39,25	0,0017	44,99	44,80	45,19	0,0010	3,02
2014	3º	38,92	38,59	39,25	0,0017	44,87	44,67	45,06	0,0010	2,98
2014	4º	39,02	38,69	39,34	0,0017	44,97	44,78	45,17	0,0010	2,81
2015	1º	38,69	38,37	39,02	0,0017	44,48	44,29	44,68	0,0010	2,85
2015	2º	38,81	38,47	39,16	0,0017	44,56	44,36	44,75	0,0010	3,11
2015	3º	39,15	38,81	39,50	0,0017	45,04	44,84	45,23	0,0010	3,07
2015	4º	38,45	38,09	38,81	0,0018	44,42	44,22	44,61	0,0010	3,35
2016	1º	38,15	37,79	38,52	0,0019	44,13	43,94	44,33	0,0010	3,40
2016	2º	38,75	38,39	39,12	0,0019	44,46	44,27	44,66	0,0010	3,39
2016	3º	38,52	38,16	38,87	0,0018	44,29	44,09	44,49	0,0010	3,18
2016	4º	38,73	38,36	39,11	0,0019	44,34	44,14	44,54	0,0010	3,47
2017	1º	38,87	38,52	39,23	0,0018	44,14	43,94	44,34	0,0010	3,18
2017	2º	40,03	39,68	40,39	0,0018	45,30	45,10	45,51	0,0010	3,17
2017	3º	40,57	40,20	40,94	0,0019	45,62	45,42	45,82	0,0010	3,43
2017	4º	40,82	40,46	41,19	0,0019	45,78	45,58	45,98	0,0010	3,35

Fonte: Elaboração própria com base na PNAD Contínua 2012-2017

Tabela 7: Resultados para Taxa de Emprego Formal (Brasil – PNADC 2012-2017)

Ano	Trimestre	Estimativa Plano amostral PNADC				Estimativa Plano amostral sob AAS				EPA estimado
		Média (%)	Limite Inferior (%)	Limite Superior (%)	Erro padrão	Média (%)	Limite Inferior (%)	Limite Superior (%)	Erro padrão	
2012	1º	58,30	57,96	58,63	0,0017	53,11	52,91	53,31	0,0010	2,86
2012	2º	58,25	57,91	58,60	0,0017	53,09	52,89	53,28	0,0010	3,02
2012	3º	58,63	58,30	58,96	0,0017	53,27	53,07	53,47	0,0010	2,84
2012	4º	58,82	58,48	59,16	0,0017	53,25	53,05	53,45	0,0010	2,93
2013	1º	59,23	58,89	59,57	0,0017	53,85	53,65	54,04	0,0010	2,94
2013	2º	59,33	58,99	59,66	0,0017	53,91	53,71	54,10	0,0010	2,91
2013	3º	59,60	59,27	59,93	0,0017	54,03	53,83	54,22	0,0010	2,80
2013	4º	59,72	59,39	60,05	0,0017	54,05	53,86	54,25	0,0010	2,86
2014	1º	60,79	60,45	61,12	0,0017	54,76	54,57	54,96	0,0010	2,91
2014	2º	61,09	60,75	61,42	0,0017	55,01	54,81	55,20	0,0010	3,02
2014	3º	61,08	60,75	61,41	0,0017	55,13	54,94	55,33	0,0010	2,98
2014	4º	60,98	60,66	61,31	0,0017	55,03	54,83	55,22	0,0010	2,81
2015	1º	61,31	60,98	61,63	0,0017	55,52	55,32	55,71	0,0010	2,85
2015	2º	61,19	60,84	61,53	0,0017	55,44	55,25	55,64	0,0010	3,11
2015	3º	60,85	60,50	61,19	0,0017	54,96	54,77	55,16	0,0010	3,07
2015	4º	61,55	61,19	61,91	0,0018	55,58	55,39	55,78	0,0010	3,35
2016	1º	61,85	61,48	62,21	0,0019	55,87	55,67	56,06	0,0010	3,40
2016	2º	61,25	60,88	61,61	0,0019	55,54	55,34	55,73	0,0010	3,39
2016	3º	61,48	61,13	61,84	0,0018	55,71	55,51	55,91	0,0010	3,18
2016	4º	61,27	60,89	61,64	0,0019	55,66	55,46	55,86	0,0010	3,47
2017	1º	61,13	60,77	61,48	0,0018	55,86	55,66	56,06	0,0010	3,18
2017	2º	59,97	59,61	60,32	0,0018	54,70	54,49	54,90	0,0010	3,17
2017	3º	59,43	59,06	59,80	0,0019	54,38	54,18	54,58	0,0010	3,43
2017	4º	59,18	58,81	59,54	0,0019	54,22	54,02	54,42	0,0010	3,35

Fonte: Elaboração própria com base na PNAD Contínua 2012-2017

5 - CONSIDERAÇÕES FINAIS

No estudo realizado apresentamos estimativas de várias taxas que geram um olhar sobre a situação do mercado de trabalho no Brasil, no entanto, não temos o objetivo de esclarecer os motivos econômicos que as referidas taxas chegaram a esses números e sim entender os impactos em desconsiderar o efeito do plano amostral em um pesquisa de grande escala como a PNAD Contínua.

Nessa monografia o estudo acompanhou a PNAD-C trimestralmente do período de 2012 a 2017 e como demonstrou nas Tabelas anteriores, vemos de forma constante que o plano amostral mais ingênuo em relação ao plano amostral complexo, possui as estimativas pontuais sempre como valores notoriamente diferentes e erros-padrão sempre menores ocasionando intervalos de confiança mais estreitos (SILVA et al, 2002, p. 3), as estimativas pontuais da população são influenciadas pelos pesos das observações. Já estimativas de variância, desvio-padrão e as estimativas de parâmetros para ajustes de alguns modelos são influenciadas pela estratificação, conglomeração e pesos. Sendo assim, ignorar tais elementos, pode levar a produção de resultados incorretos, tanto para as estimativas pontuais como para os respectivos desvios-padrões e níveis de significância. De modo geral, temos situações em que ao ignorar tais elementos do plano amostral as tomadas de decisões realizadas pelos agentes públicos, que afetam às políticas públicas de um país diretamente, podem ser equivocadas e somado a isso, deturpar a real interpretação de uma série histórica, por exemplo.

Também foi verificado o efeito do plano amostral (EPA) e constatamos em todos os trimestres para a taxa de participação na força de trabalho, nível de ocupação, taxa de ocupação, nível de desocupação, taxa de desocupação, taxa de formalidade e taxa de informalidade valores superiores a 1, principalmente para os últimos trimestres de 2016 e 2017, demonstrando que os efeitos do plano amostral não podem de forma alguma ser ignorados. Além disso, estimativas de variância baseadas na hipótese de AAS subestimam as variâncias corretas (SILVA, 1998, p.49). Comprovando a importância de considerar o plano amostral verdadeiro ao estimar variâncias e desvios padrões associados às estimativas pontuais.

Portanto, fica claro que ignorar os aspectos de um plano amostral de pesquisas como a complexidade da PNADC produzirá medidas incertas e haverá interpretações erradas e viciadas, ou seja, as análises dos dados da PNAD Contínua devem sempre considerar os efeitos do plano amostral.

Propomos para futuras análises a aplicação de estudos que envolvam a taxa de Subutilização da Força de trabalho que tem como proposta complementar o monitoramento do mercado de trabalho e em conjunto com a medida de desocupação, fornecem estimativas com relação a demanda por trabalho em ocupação. A estes estudos é interessante acrescentar debates e reflexões mais profunda sobre a taxa de informalidade do mercado de trabalho do Brasil, observando o grande impacto que gera sobre a forma como ocorre a arrecadação da previdência social no país (Menezes e Dedecca, 2012, p. 39).

E ainda, propomos uma ampliação dos estudos do efeito do plano amostral, dado que, o EPA, fornece informações para o apoio ao planejamento de novas pesquisas amostrais, mas no que se refere a análises das amostras que já foram selecionadas ele perde sua importância, devendo ser aplicado o conceito proposto por Skinner, Holt e Smith (1989, p. 24), o EPA ampliado, também conhecido como *misspecification effect (meff)*.

Vale lembrar, que hoje dispomos de uma série de pacotes estatístico e estes possuem recursos computacionais suficientes para realizar inúmeras análises com o vasto banco de informações disponibilizado pela PNAD-C, cabendo apenas, a correta implementação do plano amostral da pesquisa. Com essa monografia esperamos dar um vislumbre dos métodos de amostragem e ressaltar a atenção extra que deve ser dada ao aplicar estes métodos em análises de dados, principalmente em grandes pesquisas, pois a amostragem aplicada no seu rigor é uma poderosa ferramenta no planejamento amostral, na coleta e na análise de dados.

REFERÊNCIAS

BOLFARINE, H.; BUSSAB, W. O. Elementos de Amostragem. São Paulo, Editora Edgard Blücher, 2004.

CASELLA, G; BERGER, R.L. Statistical Inference. Duxbury, 2nd ed., 2001.

COCHRAN W. G. Sampling techniques. 3. ed. New York, USA: John Wiley & Sons, 1977.

HORVITZ D. G.; THOMPSON D. J. A Generalization of Sampling Without Replacement from a Finite Universe Journal of the American Statistical Association p. 663-685., 1952.

IBGE (Instituto Brasileiro de Geografia e Estatística). Sistema Integrado de Pesquisas Domiciliares - SIPD. Coordenação de Trabalho e Rendimento. - Rio de Janeiro (Textos para discussão. Diretoria de Pesquisas, ISSN 1518-675X; n. 24), 2007.

_____. Resolution concerning statistics of work, employment and labour underutilization. In: The Nineteenth International Conference of Labour Statisticians, 2013.

_____. CNEFE: cadastro nacional de endereços para fins estatísticos: atualização do setor: manual do atualizador. Rio de Janeiro, 2013.

_____. Pesquisa Nacional por Amostra de Domicílios Contínua Notas Metodológicas. Ministério do Planejamento, Orçamento e Gestão Instituto Brasileiro de Geografia e Estatística – IBGE, Diretoria de Pesquisas Coordenação de Trabalho e Rendimento, v.1. Rio de Janeiro, 2014.

_____. PNAD Contínua. Dicionário e Input. Disponível em:
https://www.ibge.gov.br/estatisticas-novoportal/sociais/habitacao/17270-pnad-continua.html?=&t=downloads/Trimestral/Microdados/Documentacao/Dicionario_e_input.zip. Acesso em: 07 abr 2019.

_____. Síntese de indicadores sociais: uma análise das condições de vida da população brasileira. Coordenação de População e Indicadores Sociais. Rio de Janeiro, 2016.

_____. PNAD Contínua. Microdados. Disponível em:
<https://www.ibge.gov.br/estatisticas-novoportal/sociais/habitacao/17270-pnad-continua.html?=&t=downloads/Trimestral/Microdados/> . Acesso em: 07 abr 2019.

IPEA (Instituto de Pesquisa Econômica Aplicada). Mercado de trabalho: conjuntura e análise. Ministério do Trabalho. v.1. Brasília, 2017.

KISH, L. Survey sampling. John Wiley and Sons, Inc., New York, 1965.

MENEZES W. F.; DEDECCA C. S. A informalidade no mercado de Trabalho Brasileiro: Rendimentos e principais características. v. 6, n. 2, p. 19 dez., 2012.

PESSOA, D. G. C.; SILVA, P. L. N.; DUARTE, R. P. N. Análise estatística de dados de pesquisas por amostragem: problemas no uso de pacotes padrões. Revista Brasileira de Estatística, v. 59, p. 53-76, 1997.

QUINTSLR, M. M. M. et al. Sistema Integrado de Pesquisas Domiciliares (SIPD). Rio de Janeiro: Coordenação de Trabalho e Rendimento. Diretoria de Pesquisas, Instituto Brasileiro de Geografia e Estatística, 2007. (Texto para Discussão n. 24)

ROSS, S. Probabilidade. Um curso moderno com aplicações. Bookman. Ed.8, p. 463, 2010.

SKINNER C.J., HOLT D. e SMITH T.M.F. (eds) Analysis of Complex Surveys, Chichester, Wiley, p. 23-57, 1989.

SILVA, P. L. N.; PESSOA, D. G. C. Análise de Dados Amostrais Complexos, 1998.

SILVA, P. L. N.; PESSOA, D. G. C.; LILA, M. F. Análise Estatística de Dados da PNAD: Incorporando a Estrutura do Plano Amostral. Boletim ABRASCO, v. 7, p. 659-670, 2002.

StataCorp. Stata: Release 13. Survey Data Reference Manual. College Station, TX: StataCorp LLC, 2013.

_____. Stata: Release 15. Statistical Software. College Station, TX: StataCorp LLC, 2017.

VIEIRA, M. D. T.; SILVA, P. L. N. Notas de aula de Amostragem. Juiz de Fora, Universidade Federal de Juiz de Fora, 2015.

WENTWORTH G, SMITH D. E., HARPER H. D. Fundamentals of practical mathematics - Ratio and Proportion. Ginn and Co. pp. 55ff, 1922.

WEST, B. T., BERGLUND, P., & HEERINGA, S. G. A Closer Examination of Subpopulation Analysis of Complex-Sample Survey Data. The Stata Journal, 8(4), 520–531, 2008.

APÊNDICE A – CÓDIGO UTILIZADOS

Os códigos aqui apresentados foram implementados no software *Stata 15* (STATA CORP, 2017) de propriedade da Faculdade de Economia – UFJF. E foram estruturados para estimar o efeito do plano amostral da Pesquisa Nacional por Amostras de Domicílios Contínua do período de 2012 a 2017 com relação a um plano amostral sob AAS. A programação foi inteiramente realizada em um computador *Intel® Core™ i7* com memória *RAM DDR3 16GB* e *HD* de 1 *Terabyte* rodando em sistema operacional *Microsoft Windows 7* de propriedade da Faculdade de Economia da UFJF no Laboratório de Estudos Econômicos (ECONS).

Além disso, a quem interessar, o código é flexível e permite sua execução para intervalos de anos de interesse, por exemplo do ano de 2012 ao ano em que ocorra a leitura desse trabalho, desde que tenha a disponibilidade do banco de dados para os anos em estudo, no formato “.dta”. E ainda, optamos pela criação de postfiles, pois dessa forma termos os resultados salvos em um arquivo “.dta” à parte, permitindo uma série de manipulações, tais como gráficos e ganhado em agilidade. já que não será necessário rodar o programa toda vez que quiser obter as estimações.

Apêndice A.1 – Taxa de Participação na Força de Trabalho.

```
*****
***** PNADC - CÁLCULO DA TAXA DE PARTICIPAÇÃO BRASIL (2012-2017) *****
*****

set more off

clear all

cd C:\local_oude_será_criado_o_postfile\"

** Criando um "postfile" nomeado "taxa_p" que conterà TxMed, LInf e
LSup.
postfile taxa_p Ano Trimestre LInfl TxMed1 LSup1 ERRO1 Inf2 TxMed2
LSup2 ERRO2 EPA using "tx_part_br", replace

** Criando uma estrutura recursiva "for" para varrer cada pasta com os
anos da PNADC.
forval ano = 2012/2017{
```

```

** Criando uma estrutura recursiva "for" para varrer cada uma das
pastas com os 4 trimestres da PNADC.
forval trim = 1/4{

** Abrindo o banco de dados
cap use "C:\Banco_de_dados_PNADC\\`ano'\PNADC_0`trim'\ano'.dta" ,
clear

if _rc != 0{
exit
}
cap toupper

** Calculando pessoas maiores de 14 anos.
* V2009: Idade do morador na data de referênciã.
gen pia= V2009>=14 & V2009<.

** Design da pesquisa (PNADC)
svyset UPA [pw=V1027], strata(ESTRATO) single(centered) posts(POSEST)
postw(V1029) clear

** Cálculo da taxa de participação na força de trabalho.
* Considerar apenas pessoas na força de trabalho.
cap drop subpopFT subpopLF
cap gen subpopFT= VD4001==1 // considerar 1. Pessoas na força de
trabalho
cap gen subpopLF= VD4001!=. // corrigir missing

* Cálculo da taxa, percentual de pessoas na força de trabalho em
relação de pessoas em idade de trabalhar.
svy, subpop(subpopLF): ratio subpopFT pia, cformat(%9,4fc)

* Criando a matriz "a" onde será atribuído os valores de interesse
(Média, Limite Sup e Limite Inf).
mat a = r(table)

** Design sob AAS
svyset _n

* Cálculo da taxa de desemprego sob AAS

```

```

svy, subpop(subpopLF): ratio subpopFT pia, cformat(%9,4fc)

* Criando a matriz "b" onde será atribuído os valores de interesse
(Média, Limite Sup e Limite Inf).
mat b = r(table)

* Cálculo do EPA estimado.
gen efeito = (a[2,1])^2/ (b[2,1])^2

*Criando "taxa_p" objeto que acumula todos os valores e estes são
multiplicados por 100 para serem expressos em porcentagem, exceto
erros e efeito.*
post taxa_p (`ano') (`trim') (a[5,1]*100) (a[1,1]*100) (a[6,1]*100)
(a[2,1]) (b[5,1]*100) (b[1,1]*100) (b[6,1]*100) (b[2,1]) (efeito)
}
}
postclose taxa_p /* Finalizando a criação postfile. */

*****
*****
*****

```

Apêndice A.2 – Nível de Ocupação.

```

*****
***** PNADC - CÁLCULO NÍVEL DE OCUPAÇÃO BRASIL (2012-2017) *****
*****

set more off

clear all

cd C:\local_oude_será_criado_o_postfile\"

** Criando um "postfile" nomeado "nivel_o" que conterà TxMed, LInf,
LSup e EPA.
postfile nivel_o Ano Trimestre LInf1 TxMed1 LSup1 ERRO1 Inf2 TxMed2
LSup2 ERRO2 EPA using "nv_ocup_br", replace

** Criando uma estrutura recursiva "for" para varrer cada pasta com os
anos da PNADC.

```

```

forval ano = 2012/2017{

** Criando uma estrutura recursiva "for" para varrer cada uma das
pastas com os 4 trimestres da PNADC.
forval trim = 1/4{

** Abrindo o banco de dados
"C:\Banco_de_dados_PNADC\\`ano'\PNADC_0`trim'\ano'.dta" , clear

if _rc != 0{
exit
}
cap toupper
** Calculando pessoas maiores de 14 anos.
* V2009: Idade do morador na data de referência.
gen pia= V2009>=14 & V2009<.

** Individuals que estão empregados.
gen ocup_c =VD4002==1 if pia==1

** Design da pesquisa (PNADC)
svyset UPA [pw=V1027], strata(ESTRATO) single(centered) posts(POSEST)
postw(V1029) clear

** Cálculo nível da Ocupação.
* Considerar apenas pessoas na força de trabalho.
gen subpopFT = VD4001==1

* Cálculo do nível da ocupação, percentual de pessoas ocupadas na
semana de referência em relação às pessoas em idade de trabalhar.
svy, subpop(pia): ratio ocup_c pia, cformat(%9,4fc)

* Criando a matriz "a" onde será atribuído os valores de interesse
(Média, Limite Sup e Limite Inf).
mat a = r(table)

** Design sob AAS
svyset _n

* Cálculo da taxa de desemprego sob AAS
svy, subpop(pia): ratio ocup_c pia, cformat(%9,4fc)

```

```

* Criando a matriz "b" onde será atribuído os valores de interesse
(Média, Limite Sup e Limite Inf).
mat b = r(table)

* Cálculo do EPA estimado.
gen efeito = (a[2,1])^2/ (b[2,1])^2

*Criando "nivel_o" objeto que acumula todos os valores e estes são
multiplicados por 100 para serem expressos em porcentagem, exceto
erros e efeito.*
post nivel_o (`ano') (`trim') (a[5,1]*100) (a[1,1]*100) (a[6,1]*100)
(a[2,1]) (b[5,1]*100) (b[1,1]*100) (b[6,1]*100) (b[2,1]) (efeito)
}
}
postclose nivel_o /* Finalizando a criação postfile. */

```

```

*****
*****
*****

```

Apêndice A.3 – Taxa de Ocupação.

```

*****
***** PNADC - CÁLCULO TAXA DE OCUPAÇÃO BRASIL (2012-2017) *****
*****

```

```
set more off
```

```
clear all
```

```
cd C:\local_oude_será_criado_o_postfile\
```

```
** Criando um "postfile" nomeado "taxa_o" que conterà TxMed, LInf,
LSup e EPA.
```

```
postfile taxa_o Ano Trimestre LInf1 TxMed1 LSup1 ERRO1 Inf2 TxMed2
LSup2 ERRO2 EPA using "tx_ocup_br", replace
```

```

** Criando uma estrutura recursiva "for" para varrer cada pasta com os
anos da PNADC.
forval ano = 2012/2017{

** Criando uma estrutura recursiva "for" para varrer cada uma das
pastas com os 4 trimestres da PNADC.
forval trim = 1/4{

** Abrindo o banco de dados
cap use C:\Banco_de_dados_PNADC\\ano'\PNADC_0`trim`ano'.dta" , clear

if _rc != 0{
exit
}
cap toupper

** Individuals que estão empregados.
gen ocup_c =VD4002==1 & VD4002 != .

** Cálculo da taxa de desemprego.
* Considerar apenas pessoas na força de trabalho.
gen subpopFT = VD4001==1 & VD4001 != .

** Design da pesquisa (PNADC)
svyset UPA [pw=V1027], strata(ESTRATO) single(centered) posts(POSEST)
postw(V1029) clear

* Cálculo da taxa de desemprego, porção de desempregados da força
total de trabalho.
svy, subpop(subpopFT): ratio ocup_c subpopFT, cformat(%9,4fc)12

* Criando a matriz "a" onde será atribuído os valores de interesse
(Média, Limite Sup e Limite Inf).
mat a = r(table)

** Design sob AAS
svyset _n

* Cálculo da taxa de desemprego sob AAS
svy, subpop(subpopFT): ratio ocup_c subpopFT, cformat(%9,4fc)

```

```

* Criando a matriz "b" onde será atribuído os valores de interesse
(Média, Limite Sup e Limite Inf).
mat b = r(table)

* Cálculo do EPA estimado.
gen efeito = (a[2,1])^2/(b[2,1])^2

*Criando "taxa_o" objeto que acumula todos os valores e estes são
multiplicados por 100 para serem expressos em porcentagem, exceto
erros e efeito.*
post taxa_o (`ano') (`trim') (a[5,1]*100) (a[1,1]*100) (a[6,1]*100)
(a[2,1]) (b[5,1]*100) (b[1,1]*100) (b[6,1]*100) (b[2,1]) (efeito)
}
}
postclose taxa_o /* Finalizando a criação postfile. */

```

```

*****
*****
*****

```

Apêndice A.4 – Nível de Desocupação.

```

*****
***** PNADC - CÁLCULO DO NÍVEL DE DESOCUPAÇÃO BRASIL (2012-2017) *****
*****

```

```
set more off
```

```
clear all
```

```
cd "C:\local_oude_será_criado_o_postfile\"
```

```
** Criando um "postfile" nomeado "nivel_d" que conterà TxMed's,
LInf's, LSup's, erro's e EPA.
```

```
postfile nivel_d Ano Trimestre LInf1 TxMed1 LSup1 ERRO1 Inf2 TxMed2
LSup2 ERRO2 EPA using "nv_des_br", replace
```

```
** Criando uma estrutura recursiva "for" para varrer cada pasta com os
anos da PNADC.
```

```
forval ano = 2012/2017{
```

```

** Criando uma estrutura recursiva "for" para varrer cada uma das
pastas com os 4 trimestres da PNADC.
forval trim = 1/4{

** Abrindo o banco de dados
cap use "C:\Banco_de_dados_PNADC\\`ano'\PNADC_0`trim`ano'.dta" ,
clear

if _rc != 0{
exit
}
cap toupper

** Calculando pessoas maiores de 14 anos.
* V2009: Idade do morador na data de referência.
gen pia= V2009>=14 & V2009<.

** Indivíduos que estão desempregados.
gen desocup = VD4002==2 if pia==1

** Design da pesquisa (PNADC)
svyset UPA [pw=V1027], strata(ESTRATO) single(centered) posts(POSEST)
postw(V1029) clear

** Cálculo Nível da Ocupação.
* Considerar apenas pessoas na força de trabalho.

* Cálculo do Nível da Ocupação, percentual de pessoas ocupadas na
semana de referência em relação às pessoas em idade de trabalhar.
svy, subpop(pia): ratio desocup pia, cformat(%9,4fc)

* Criando a matriz "a" onde será atribuído os valores de interesse
(Média, Limite Sup e Limite Inf).
mat a = r(table)

** Design sob AAS
svyset _n

* Cálculo da taxa de desemprego sob AAS
svy, subpop(pia): ratio desocup pia, cformat(%9,4fc)

```

```

* Criando a matriz "b" onde será atribuído os valores de interesse
(Média, Limite Sup e Limite Inf).
mat b = r(table)

* Cálculo do EPA estimado.
gen efeito = (a[2,1])^2/(b[2,1])^2

*Criando "nivel_d" objeto que acumula todos os valores e estes são
multiplicados por 100 para serem expressos em porcentagem, exceto
erros e efeito.*
post nivel_d (`ano') (`trim') (a[5,1]*100) (a[1,1]*100) (a[6,1]*100)
(a[2,1]) (b[5,1]*100) (b[1,1]*100) (b[6,1]*100) (b[2,1]) (efeito)
}
}
postclose nivel_d /* Finalizando a criação postfile. */

*****
*****
*****

```

Apêndice A.5 – Taxa de Desocupação.

```

*****
***** PNADC - CÁLCULO DA TAXA DE DESEMPREGO BRASIL (2012-2017) *****
*****

set more off

clear all

cd "C:\local_oude_será_criado_o_postfile\"

** Criando um "postfile" nomeado "taxa_d" que conterà TxMed, LInf,
LSup, erro e EPA.
postfile taxa_d Ano Trimestre LInf1 TxMed1 LSup1 ERRO1 Inf2 TxMed2
LSup2 ERRO2 EPA using "tx_des_br", replace

** Criando uma estrutura recursiva "for" para varrer cada pasta com os
anos da PNADC.
forval ano = 2012/2017{

```

```

** Criando uma estrutura recursiva "for" para varrer cada uma das
pastas com os 4 trimestres da PNADC.
forval trim = 1/4{

** Abrindo o banco de dados
cap use "C:\Banco_de_dados_PNADC\\`ano'\PNADC_0`trim`ano'.dta" ,
clear
if _rc != 0{
exit
}
cap toupper

gen desocup=VD4002==2 & VD4002 !=.

* Considerar apenas pessoas na força de trabalho.
cap drop subpopFT subpopLF
cap gen subpopFT= VD4001==1 // considerar 1. Pessoas na força de
trabalho
cap gen subpopLF= VD4001!=. // corrigir missing

** Design da pesquisa (PNADC)
svyset UPA [pw=V1027], strata(ESTRATO) single(centered) posts(POSEST)
postw(V1029) clear

* Cálculo da taxa de desemprego, proporção de desempregados na força
total de trabalho.
svy, subpop(subpopLF): ratio desocup subpopFT

* Criando a matriz "a" onde será atribuído os valores de interesse
(Média, Limite Sup e Limite Inf).
mat a = r(table)

** Design sob AAS
svyset _n

* Cálculo da taxa de desemprego sob AAS
svy, subpop(subpopLF): ratio desocup subpopFT

* Criando a matriz "b" onde será atribuído os valores de interesse
(Média, Limite Sup e Limite Inf).
mat b = r(table)

```

```

* Cálculo do EPA estimado.
gen efeito = (a[2,1])^2/(b[2,1])^2

*Criando "taxa_d" objeto que acumula todos os valores e estes são
multiplicados por 100 para serem expressos em porcentagem, exceto
erros e efeito.*
post taxa_d (`ano') (`trim') (a[5,1]*100) (a[1,1]*100) (a[6,1]*100)
(a[2,1]) (b[5,1]*100) (b[1,1]*100) (b[6,1]*100) (b[2,1]) (efeito)
}
}
postclose taxa_d /* Finalizando a criação do postfile. */

*****
*****
*****

```

Apêndice A.6 – Taxa de Informalidade.

```

*****
**** PNADC - CÁLCULO DA TAXA DE INFORMALIDADE BRASIL (2012-2017) ****
*****
set more off

clear all

cd "C:\local_onde_será_criado_o_postfile\"

** Criando um "postfile" nomeado "taxa_i" que conterà TxMed, LInf e
LSup .
postfile taxa_i Ano Trimestre LInf1 TxMed1 LSup1 ERRO1 Inf2 TxMed2
LSup2 ERRO2 EPA using "tx_trabinf_br", replace

** Criando uma estrutura recursiva "for" para varrer cada pasta com os
anos da PNADC.
forval ano = 2012/2017{

** Criando uma estrutura recursiva "for" para varrer cada uma das
pastas com os 4 trimestres da PNADC.
forval trim = 1/4{

```

```

** Abrindo o banco de dados
cap use "C:\Banco_de_dados_PNADC\\`ano'\PNADC_0`trim'\ano'.dta" ,
clear

if _rc != 0{
exit
}
cap toupper
** Calculando pessoas maiores de 14 anos.
* V2009: Idade do morador na data de referência.
gen pia= V2009>=14 & V2009<.

** Design da pesquisa (PNADC)

svyset UPA [pw=V1027], strata(ESTRATO) single(centered) posts(POSEST)
postw(V1029) clear

** Trabalho formal: Trabalho com carteira assinada, incluindo os
trabalhadores domésticos; militares e funcionários públicos
estatutários, bem como os empregadores e
**trabalhadores por conta própria que contribuem para a previdência
social.

* Trabalhador setor privado com carteira assinada.
gen tpcc=VD4009==1

* Trabalhador domésticos com carteira assinada.
gen tdcc=VD4009==3

* Trabalhador do setor público com carteira assinada.
gen tpucc=VD4009==5

* Militares e funcionários públicos estatutários
gen mse=VD4009==7

* Empregador que contribuem para a previdência social
gen empcp=VD4009==8 if VD4012==1

* Trabalhador por conta própria que contribuem para a previdência
social
gen cpcp=VD4009==9 if VD4012==1

```

```

** Trabalho informal: Trabalho sem carteira assinada, incluindo os
trabalhadores domésticos, empregadores e trabalhadores conta própria
que não contribuem para a

**previdência social, trabalhadores não remunerados, bem como os
trabalhadores na
**produção para o próprio consumo e na construção para o próprio uso.

* Trabalhador setor privado sem carteira assinada.
gen tpssc=VD4009==2

* Trabalhador domésticos sem carteira assinada.
gen tdsc=VD4009==4

* Trabalhador do setor público sem carteira assinada.
gen tpusc=VD4009==6

* Empregador que não contribuem para a previdência social
gen empncp=VD4009==8 if VD4012==2

* Trabalhador por conta própria que não contribuem para a previdência
social
gen cpncp=VD4009==9 if VD4012==2

* Trabalhador familiar auxiliar
gen tf=VD4009==10

** Trabalhadores formais
egen formal = rowtotal (tpcc tdcc tpucc mse empcc cpcc)

* Trabalhadores informais
egen informal = rowtotal (tpssc tdsc tpusc empncp cpncp tf)

** Trabalhadores ocupados (formais + informais)
egen trabocup = rowtotal (formal informal)

* Cálculo do Taxa de Informalidade, percentual de pessoas ocupadas na
semana de referência, mas que não realizaram nenhuma contribuição para
previdência social.
svy, subpop(trabocup): ratio informal trabocup, cformat(%9,4fc)

```

```

* Criando a matriz "a" onde será atribuído os valores de interesse
(Média, Limite Sup e Limite Inf).
mat a = r(table)

** Design sob AAS
svyset _n

* Cálculo da taxa de informalidade sob AAS
svy, subpop(trabocup): ratio informal trabocup, cformat(%9,4fc)

* Criando a matriz "b" onde será atribuído os valores de interesse
(Média, Limite Sup e Limite Inf).
mat b = r(table)

* Cálculo do EPA estimado.
gen efeito = (a[2,1])^2/ (b[2,1])^2

*Criando "taxa_i" objeto que acumula todos os valores e estes são
multiplicados por 100 para serem expressos em porcentagem, exceto
erros e efeito. *
post taxa_i (`ano') (`trim') (a[5,1]*100) (a[1,1]*100) (a[6,1]*100)
(a[2,1]) (b[5,1]*100) (b[1,1]*100) (b[6,1]*100) (b[2,1]) (efeito)
}
}
postclose taxa_i /* Finalizando a criação postfile. */

```

Apêndice A.7 – Taxa de Emprego Formal.

```

*****
**** PNADC - CÁLCULO DA TAXA DE EMPREGO FORMAL BRASIL (2012-2017) ****
*****

set more off

clear all

cd "C:\local_onde_será_criado_o_postfile\"

** Criando um "postfile" nomeado "taxa_f" que conterà TxMed, LInf e
LSup .
postfile taxa_f Ano Trimestre LInf1 TxMed1 LSup1 ERRO1 Inf2 TxMed2
LSup2 ERRO2 EPA using "tx_trabfor_br", replace

```

```

** Criando uma estrutura recursiva "for" para varrer cada pasta com os
anos da PNADC.
forval ano = 2012/2017{

** Criando uma estrutura recursiva "for" para varrer cada uma das
pastas com os 4 trimestres da PNADC.
forval trim = 1/4{

** Abrindo o banco de dados
cap use "C:\Banco_de_dados_PNADC\\`ano'\PNADC_0`trim'\ano'.dta" ,
clear

if _rc != 0{
    exit
}
cap toupper
** Calculando pessoas maiores de 14 anos.
* V2009: Idade do morador na data de referência.
gen pia= V2009>=14 & V2009<.

** Design da pesquisa (PNADC)
svyset UPA [pw=V1027], strata(ESTRATO) single(centered) posts(POSEST)
postw(V1029) clear

** Trabalho formal: Trabalho com carteira assinada, incluindo os
trabalhadores domésticos; militares e funcionários públicos
estatutários, bem como os empregadores e
**trabalhadores por conta própria que contribuem para a previdência
social.

* Trabalhador setor privado com carteira assinada.
gen tpcc=VD4009==1

* Trabalhador domésticos com carteira assinada.
gen tdcc=VD4009==3

* Trabalhador do setor público com carteira assinada.
gen tpucc=VD4009==5

* Militares e funcionários públicos estatutários

```

gen mse=VD4009==7

* Empregador que contribuem para a previdência social

gen empcp=VD4009==8 if VD4012==1

* Trabalhador por conta própria que contribuem para a previdência social

gen cpcp=VD4009==9 if VD4012==1

** Trabalho informal: Trabalho sem carteira assinada, incluindo os trabalhadores domésticos, empregadores e trabalhadores conta própria que não contribuem para a

**previdência social, trabalhadores não remunerados, bem como os trabalhadores na

**produção para o próprio consumo e na construção para o próprio uso.

* Trabalhador setor privado sem carteira assinada.

gen tpsc=VD4009==2

* Trabalhador domésticos sem carteira assinada.

gen tdsc=VD4009==4

* Trabalhador do setor público sem carteira assinada.

gen tpusc=VD4009==6

* Empregador que não contribuem para a previdência social

gen empncp=VD4009==8 if VD4012==2

* Trabalhador por conta própria que não contribuem para a previdência social

gen cpncp=VD4009==9 if VD4012==2

* Trabalhador familiar auxiliar

gen tf=VD4009==10

** Trabalhadores formais

egen formal = rowtotal (tpcc tdcc tpucc mse empcp cpcp)

* Trabalhadores informais

egen informal = rowtotal (tpsc tdsc tpusc empncp cpncp tf)

```

** Trabalhadores ocupados (formais + informais)
egen trabocup = rowtotal (formal informal)

* Cálculo do Taxa de Formalidade, percentual de pessoas ocupadas na
semana de referência, que realizaram alguma contribuição para
previdência social.
svy, subpop(trabocup): ratio formal trabocup, cformat(%9,4fc)

* Criando a matriz "a" onde será atribuído os valores de interesse
(Média, Limite Sup e Limite Inf).
mat a = r(table)

** Design sob AAS
svyset _n

* Cálculo da taxa de formalidade sob AAS
svy, subpop(trabocup): ratio formal trabocup, cformat(%9,4fc)

* Criando a matriz "b" onde será atribuído os valores de interesse
(Média, Limite Sup e Limite Inf).
mat b = r(table)

* Cálculo do EPA estimado.
gen efeito = (a[2,1])^2/ (b[2,1])^2

*Criando "taxa_f" objeto que acumula todos os valores e estes são
multiplicados por 100 para serem expressos em porcentagem, exceto
erros e efeito. *
post taxa_f (`ano') (`trim') (a[5,1]*100) (a[1,1]*100) (a[6,1]*100)
(a[2,1]) (b[5,1]*100) (b[1,1]*100) (b[6,1]*100) (b[2,1]) (efeito)
}
}
postclose taxa_f /* Finalizando a criação postfile. */

```