**Orientações para análises de dados da Pesquisa Nacional de Saúde:**

Assim como os demais inquéritos realizados pelo Departamento de Análise Epidemiológica e Vigilância de Doenças Não Transmissíveis, [Vigitel](https://www.gov.br/saude/pt-br/composicao/svsa/inqueritos-de-saude/vigitel) e [PeNSE](https://www.gov.br/saude/pt-br/composicao/svsa/inqueritos-de-saude/pense), as edições da [Pesquisa Nacional de Saúde](https://www.gov.br/saude/pt-br/composicao/svsa/inqueritos-de-saude/pns) devem considerar os fatores de ponderação e o desenho amostral para a análise de dados.

Para isso, é necessário o uso de programas estatísticos como o Epi Info, R, Stata, SPSS, SAS, entre outros, que possuem módulos de análise específicos para a utilização de fatores de ponderação, e delineamento amostral.

# **Pesquisa Nacional de Saúde (PNS) - edição 2013**

## **Questionário de todos os moradores e do morador selecionado**

O questionário é composto por três blocos: (1) domiciliar, referente às características do domicílio, nos moldes do censo demográfico e da PNAD; (2) o relativo a todos os moradores do domicílio, que dá continuidade ao Suplemento Saúde da PNAD; e (3) o individual, respondido por um morador de 18 anos e mais do domicílio, selecionado com equiprobabilidade entre todos os residentes elegíveis, com enfoque às principais doenças crônicas não transmissíveis, aos estilos de vida, e ao acesso ao atendimento médico. No morador adulto selecionado, foram feitas aferições de peso, altura, circunferência da cintura e pressão arterial e foi realizada a coleta de material biológico.

Ao utilizar o banco para a análise de dados, deve-se atentar à utilização dos pesos, pois estes são diferentes para o **morador selecionado**, para todos os **moradores do domicílio,** e **exames laboratoriais**. Recomenda-se a utilização do dicionário de variáveis.

* Chaveamento:
	+ Informações de domicílio: V0001 (Unidade da Federação), V0024 (Estrato), UPA\_PNS (UPA) e V0006\_PNS (Número de ordem do domicílio na PNS).
	+ Informações de morador: V0001 (Unidade da Federação), V0024 (Estrato), UPA\_PNS (UPA), V0006\_PNS (Número de ordem do domicílio na PNS) e C00301 (Número de ordem do morador).

Exemplo de como utilizar os pesos utilizando o programa estatístico **Stata**, no **módulo survey**:

1) Para análises que consideram **todos os moradores do domicílio (blocos C ao L)**, deve ser considerado:

UPA (samplingunits): UPA\_PNS

Estrato (strata): V0024

Peso do domicílio (samplingweightvariable): V00281

2) Para análises que consideram os **moradores selecionados (blocos M ao X)**, deve ser considerado:

UPA (samplingunits): UPA\_PNS

UPA\_PNS Estrato (strata): V0024

Peso do morador selecionado (samplingweightvariable): V00291

3) Comandos adicionais necessários:

vce(linearized): Especifica que a variância deve ser calculada usando o método de linearização de Taylor, comumente utilizado para estimativas robustas em dados provenientes de amostras complexas.

singleunit(missing): Define como o Stata lida com estratos contendo apenas uma unidade amostral, tratando-os como valores ausentes (missing). Útil para evitar problemas em análises com amostragem estratificada

Exemplo de declaração de **survey para morador selecionado**:

svyset UPA\_PNS [pweight=V00291], strata(V0024) vce(linearized) singleunit(centered)

Observação: Antes da análise, verifique se essas variáveis se encontram em formato numérico no banco de dados utilizado.

**Referências:**

* Concepção: <https://www.scielo.br/j/csc/a/rysffTqrwZPZnghSq5CJHsG/?lang=pt>
* Suplemento Primeiros resultados PNS 2013: <https://www.scielo.br/j/ress/i/2015.v24n2/>

## **Exames laboratoriais**

Na PNS-2013, em uma subamostra dos indivíduos selecionados para entrevista individual, foi feita coleta de material biológico para realização de exames laboratoriais. Os exames laboratoriais realizados com as amostras de sangue foram: hemoglobina glicada; colesterol total; colesterol lipoproteína de baixa densidade (LDL); colesterol lipoproteína de alta densidade (HDL); sorologia para dengue - imunoglobulina G (IgG); hemograma série vermelha (eritograma) e série branca (leucograma); cromatografia líquida de alta eficiência (HPLC) para diagnóstico de hemoglobinopatias; e creatinina. E na urina, estimativa de excreção de potássio, sal, sódio e creatinina.

O banco e o dicionário de dados provenientes da coleta de material biológico podem ser encontrados no link: <https://www.pns.icict.fiocruz.br/bases-de-dados/>

Exemplo de como utilizar os pesos utilizando o programa estatístico **Stata**, no **módulo survey**:

1) Para análises que consideram os indivíduos que participaram da coleta de material biológico, deve ser considerado:

Peso (samplingweightvariable): peso\_lab

2) Comandos adicionais necessários:

**\_n:** representa o número da linha no conjunto de dados, funcionando como um identificador automático de cada unidade de observação. No Stata, \_n é frequentemente usado para referenciar ou indexar observações individuais sem a necessidade de uma variável identificadora específica. Nesta análise, utiliza-se \_n como identificador porque a amostra é pequena, tornando desnecessário o uso de uma variável específica para identificação das unidades amostrais.

**vce(linearized):** especifica que a variância deve ser calculada usando o método de linearização de Taylor, comumente utilizado para estimativas robustas em dados provenientes de amostras complexas.

**singleunit(missing):** Define como o Stata lida com estratos contendo apenas uma unidade amostral, tratando-os como valores ausentes (missing). Útil para evitar problemas em análises com amostragem estratificada

Exemplo de declaração de survey para o material biológico:

svyset \_n [pweight=peso\_lab], vce(linearized) singleunit(missing)

Observação: Antes da análise verifique se essas variáveis se encontram em formato numérico no banco de dados utilizado.

**Referências:**

* Suplementos exames laboratoriais PNS 2013: [https://www.scielo.br/j/rbepid/i/2019.v22suppl2/https://www.scielo.br/j/rbepid/i/2019.v22suppl2/](https://www.scielo.br/j/rbepid/i/2019.v22suppl2/)
* Nota técnica IBGE exames laboratoriais: [https://ftp.ibge.gov.br/PNS/2013/Divulgacoes/Outros/Exames/Nota\_tecnica\_Exames\_Laboratoriais\_PNS.pdfhttps://ftp.ibge.gov.br/PNS/2013/Divulgacoes/Outros/Exames/Nota\_tecnica\_Exames\_Laboratoriais\_PNS.pdf](https://ftp.ibge.gov.br/PNS/2013/Divulgacoes/Outros/Exames/Nota_tecnica_Exames_Laboratoriais_PNS.pdf)

# **Pesquisa Nacional de Saúde (PNS) - edição 2019**

## **Questionário de todos os moradores e do morador selecionado**

O questionário é composto por três blocos: (1) domiciliar, referente às características do domicílio, nos moldes do censo demográfico e da PNAD; (2) o relativo a todos os moradores do domicílio, que dá continuidade ao Suplemento Saúde da PNAD; e (3) o individual, respondido por um morador de 15 anos e mais do domicílio, selecionado com equiprobabilidade entre todos os residentes elegíveis, com enfoque às principais doenças crônicas não transmissíveis, aos estilos de vida, e ao acesso ao atendimento médico. Em 2019, em uma subamostra dos moradores selecionados, foram feitas aferições de peso e altura,

Ao utilizar o banco para análise de dados, deve-se atentar à utilização dos pesos, pois estes são diferentes para o **morador selecionado,** para todos os **moradores do domicílio e para a análise de dados antropométricos**. Recomenda-se a utilização do dicionário de variáveis, disponível em:

* Chaveamento:
	+ Informações de domicílio: V0001 (Unidade da Federação), V0024 (Estrato), UPA\_PNS (UPA) e V0006\_PNS (Número de ordem do domicílio na PNS).
	+ Informações de morador: V0001 (Unidade da Federação), V0024 (Estrato), UPA\_PNS (UPA), V0006\_PNS (Número de ordem do domicílio na PNS) e C00301 (Número de ordem do morador).

Exemplo de como utilizar os pesos utilizando o programa estatístico **Stata**, no **módulo survey**:

1) Para análises que consideram **todos os moradores do domicílio (blocos C ao L)**, deve ser considerado:

UPA (samplingunits): UPA\_PNS

Estrato (strata): V0024

Peso do domicílio (samplingweightvariable): V00281

2) Para análises que consideram os **moradores selecionados (blocos M ao X)**, deve ser considerado:

UPA (samplingunits): UPA\_PNS

UPA\_PNS Estrato (strata): V0024

Peso do morador selecionado (samplingweightvariable): V00291

3) Comandos adicionais necessários:

**vce(linearized):** Especifica que a variância deve ser calculada usando o método de linearização de Taylor, comumente utilizado para estimativas robustas em dados provenientes de amostras complexas.

**singleunit(missing):** Define como o Stata lida com estratos contendo apenas uma unidade amostral, tratando-os como valores ausentes (missing). Útil para evitar problemas em análises com amostragem estratificada

Exemplo de declaração de survey para morador selecionado:

svyset UPA\_PNS [pweight=V00291], strata(V0024) vce(linearized) singleunit(centered)

Observação: Antes da análise, verifique se essas variáveis se encontram em formato numérico no banco de dados utilizado.

## **Antropometria**

Na PNS 2019, a aferição de peso e altura foi realizada em uma subamostra das pessoas selecionadas para a entrevista individual, excluídas as gestantes.

1) Para análises que consideram os **moradores selecionados para antropometria**, deve ser considerado:

UPA (samplingunits): UPA\_PNS

UPA\_PNS Estrato (strata): V0024

Peso do morador selecionado para antropometria com calibração (samplingweightvariable): V00301

2) Comandos adicionais necessários:

**vce(linearized):** Especifica que a variância deve ser calculada usando o método de linearização de Taylor, comumente utilizado para estimativas robustas em dados provenientes de amostras complexas.

**singleunit(missing):** Define como o Stata lida com estratos contendo apenas uma unidade amostral, tratando-os como valores ausentes (missing). Útil para evitar problemas em análises com amostragem estratificada

Exemplo de declaração de survey para antropometria:

svyset UPA\_PNS [pweight=V00301], strata(V0024) vce(linearized) singleunit(centered)

Observação: Antes da análise, verifique se essas variáveis se encontram em formato numérico no banco de dados utilizado.

### **Uso de subgrupos em inquéritos com desenho amostral complexo**

Em inquéritos de base populacional com desenho amostral complexo, como a Pesquisa Nacional de Saúde (PNS), a análise estatística deve respeitar a estrutura de estratificação, conglomerados e pesos amostrais, os quais são essenciais para garantir a representatividade dos resultados em relação à população-alvo. Tais elementos são incorporados ao planejamento amostral para possibilitar a inferência estatística válida, corrigindo potenciais vieses decorrentes da seleção amostral e assegurando a estimação adequada dos erros-padrão.

Nesse contexto, a exclusão direta de registros da base de dados com o objetivo de restringir a análise a um subgrupo específico (por exemplo, excluir homens para analisar apenas mulheres) é metodologicamente inadequada, pois compromete a integridade do plano amostral. Essa prática impede a utilização correta dos pesos amostrais e compromete a estimação dos parâmetros populacionais e das variâncias associadas, podendo levar a conclusões incorretas.

A forma metodologicamente correta de realizar análises em subgrupos, em inquéritos com desenho amostral complexo, é utilizar funcionalidades específicas que respeitam esse desenho, mantendo todas as observações na amostra original e aplicando restrições apenas na etapa analítica. Isso assegura que as estimativas e suas variâncias sejam ajustadas de maneira apropriada, preservando a representatividade populacional e a validade estatística das inferências para o subgrupo de interesse.

Nos principais softwares estatísticos, isso é feito por meio de comandos que permitem definir a subpopulação sem excluir os demais casos do banco de dados, garantindo que o plano amostral completo (estratos, conglomerados e pesos) seja considerado nos cálculos.

No Stata, utiliza-se a opção subpop() dentro dos comandos da família svy, após declarar o desenho amostral com svyset. No R, pode-se usar o pacote survey, que permite definir um objeto de desenho amostral e realizar análises em subgrupos com a função subset() aplicada ao objeto de desenho.

Dessa forma, o software considera toda a amostra e os pesos correspondentes, mas restringe a estimativa ao subgrupo de mulheres, realizando os ajustes necessários para a obtenção de erros-padrão e intervalos de confiança apropriados.

**Referências:**

* Métodos: https://www.scielo.br/j/ress/a/RdbtmCHjJGt8xDW6bV3Y6JB/
* Suplemento PNS 2019 RESS: https://www.scielo.br/j/ress/i/2022.v31nspe1/

Adicionalmente, há uma biblioteca em R desenvolvida pelo IBGE para análise de dados da pesquisa. Maiores detalhes podem encontrados nos links:

* [https://cran.r-project.org/web/packages/PNSIBGE/index.htmlhttps://cran.r-project.org/web/packages/PNSIBGE/index.html](https://cran.r-project.org/web/packages/PNSIBGE/index.html)
* https://cran.r-project.org/web/packages/PNSIBGE/readme/README.html