

Análise de Dados Amostrais Complexos (MAE4003) - Projeto final

Alisson Vinicius Gonsales

2023-02-23

RESUMO E OBJETIVOS

O presente trabalho foi apresentado como projeto final para a disciplina de Análise de Dados Amostrais Complexos (MAE4003) e visa a condução de análises na base de dados da amostragem PNS - 2019, utilizando-se de técnicas de amostragem complexa através dos softwares disponíveis. As análises procuram responder algumas questões de natureza descritiva da população estudada, seus hábitos e qualidade de vida. Na seção abaixo há uma descrição sucinta dos métodos utilizados na pesquisa de amostragem do PNS e as seções posteriores tentam responder as questões levantadas.

AMOSTRAGEM PNS - 2019

As pesquisas domiciliares como a PNAD (Pesquisa Nacional de Amostra Domicílios) favorecem muito a aquisição de dados que caracterizem aspectos da qualidade de vida dos brasileiros. A necessidade de pesquisas que se dedicassem exclusivamente a mapear aspectos da saúde no Brasil levou, em 2013, a criação do PNS. A Pesquisa Nacional de Saúde(PNS) é uma pesquisa domiciliar que abrange todo o território nacional e é fundamentada em três eixos principais: o desempenho do sistema nacional de saúde; as condições de saúde da população brasileira; e a vigilância das doenças crônicas não transmissíveis e os fatores de risco associados. Os resultados da PNS são vitais para o planejamento de políticas públicas que ampliem o acesso a saúde e a qualidade de vida dos brasileiros.

O plano amostral utilizado na pesquisa é uma amostragem conglomerada de três estágios. No primeiro estágio, as unidades primárias (Setores censitários) foram subamostradas da Amostra Mestra utilizando Amostragem Aleatória Simples, com pesos de seleção previamente calculados pela Amostra Mestra.

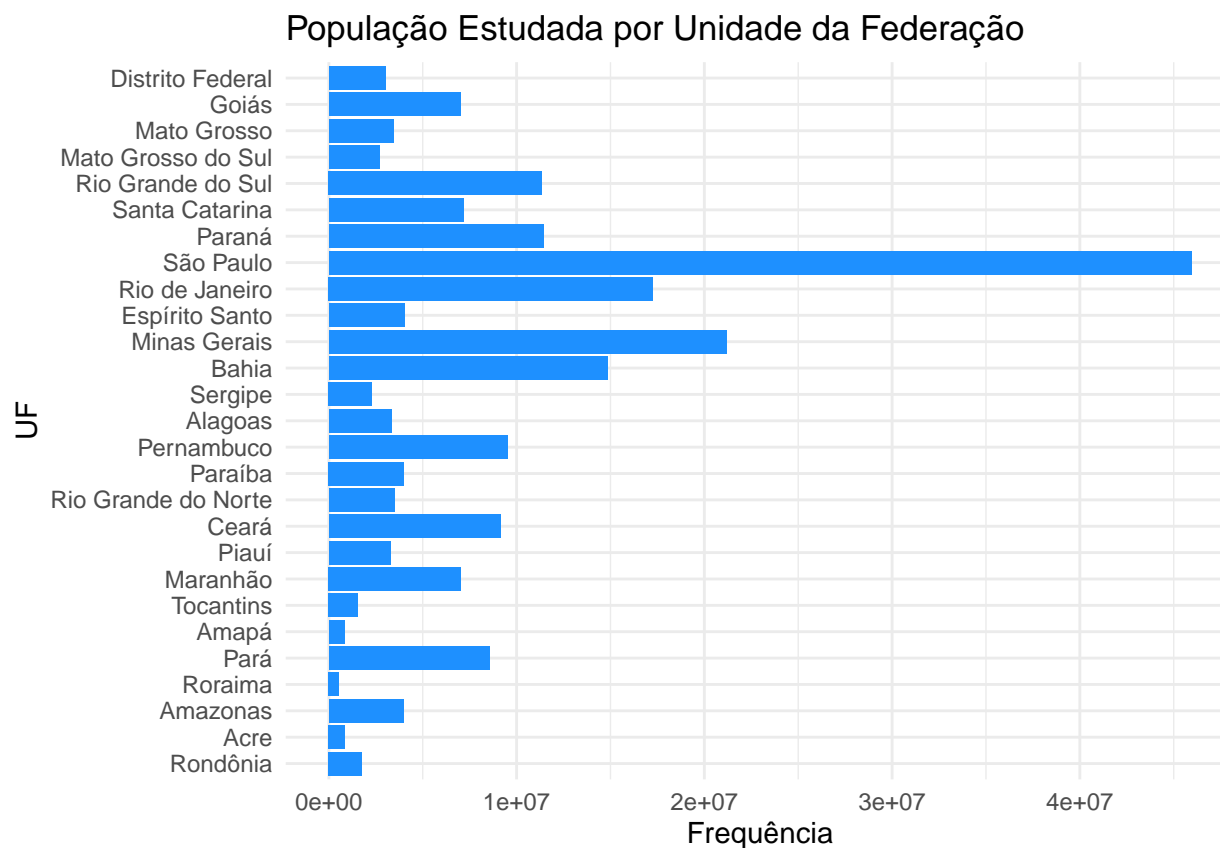
No segundo estágio, as unidades secundárias (domicílios) foram amostradas utilizando Amostragem aleatória simples, o número de domicílios amostrados por unidade primária foi inicialmente 15, com variações em alguns estados a fim de balancear as unidades primárias amostradas por Unidade da Federação. Os pesos utilizados no segundo estágio levam em conta os pesos da amostragem das unidades primárias, com ajustes para não-respostas e calibração das estimativas segundo totais populacionais calculados pela Coordenação de População e Indicadores Sociais, do IBGE.

No terceiro estágio, um morador de 15 anos ou mais em cada domicílio foi selecionado utilizando Amostragem Aleatória Simples, a partir da lista de moradores construída durante a entrevista. Os pesos utilizados nesse estágio foram calculados considerando a probabilidade de seleção dos moradores, com ajustes para não-resposta por sexo e calibração das estimativas utilizando totais populacionais por sexo e classes de idade. A entrevista continha questões comuns a todos os moradores do domicílio e questões específicas, em ambas categorias a taxa de perda de respostas ficou abaixo do planejado para o estudo. Para garantir a comparabilidade dos dados adquiridos no PNS 2019 e PNS 2013, o IBGE utilizou reponderação dos pesos da pesquisa anterior para estender a base, empregando estimadores razão com variável independente correspondente ao total populacional dos níveis geográficos, calculado segundo projeções do próprio IBGE.

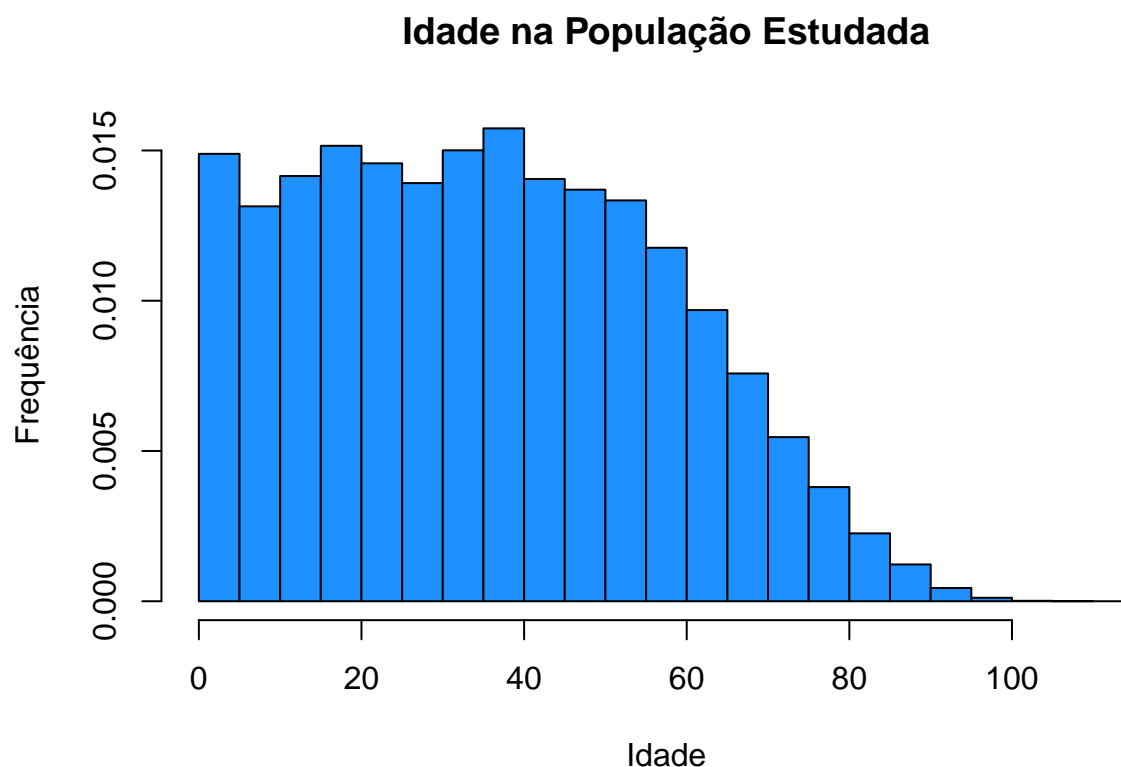
RESULTADOS

Caracterização da população-alvo do PNS 2019 de acordo com a sua demografia

Buscamos aqui caracterizar a população-alvo quanto à demografia. Abaixo vemos a distribuição da variável unidade federativa.



Vemos valores mais expressivos para São Paulo, Minas Gerais e Rio de Janeiro, com os menores valores para alguns estados do Norte. Analisando a distribuição das idades na população (figura abaixo), vemos que há frequências similares para idades na faixa dos 20-60, com frequências cada vez menores para idades acima desse valor.



Caracterizando a população estudada pelo Sexo, Cor/Etnia na tabela abaixo, podemos notar que há predominância de indivíduos que se autodeclararam brancos ou pardos, e do sexo feminino. Os intervalos de confiança calculados aderem bem ao valor médio das estimativas.

Characteristic	N = 209,589,607	95% CI
Cor ou etnia		
Branca	91,037,722 (43%)	43%, 44%
Preta	21,786,515 (10%)	10%, 11%
Amarela	1,674,691 (0.8%)	0.67%, 0.95%
Parda	94,086,142 (45%)	44%, 45%
Indígena	990,265 (0.5%)	0.41%, 0.55%
Ignorado	14,272 (<0.1%)	0.00%, 0.01%
Sexo		
Homem	100,158,109 (48%)	48%, 48%
Mulher	109,431,498 (52%)	52%, 52%

A pesquisa também traz informação sobre o nível de instrução do entrevistado, em uma variável:

Nível de instrução mais elevado alcançado (pessoas de 5 anos ou mais de idade) padronizado para o Ensino Fundamental - SISTEMA DE 9 ANOS

As categorias disponíveis são: A - Sem Instrução

B - Fundamental incompleto ou equivalente

C - Fundamental completo ou equivalente

D - Médio incompleto ou equivalente

E - Médio completo ou equivalente

F - Superior incompleto ou equivalente

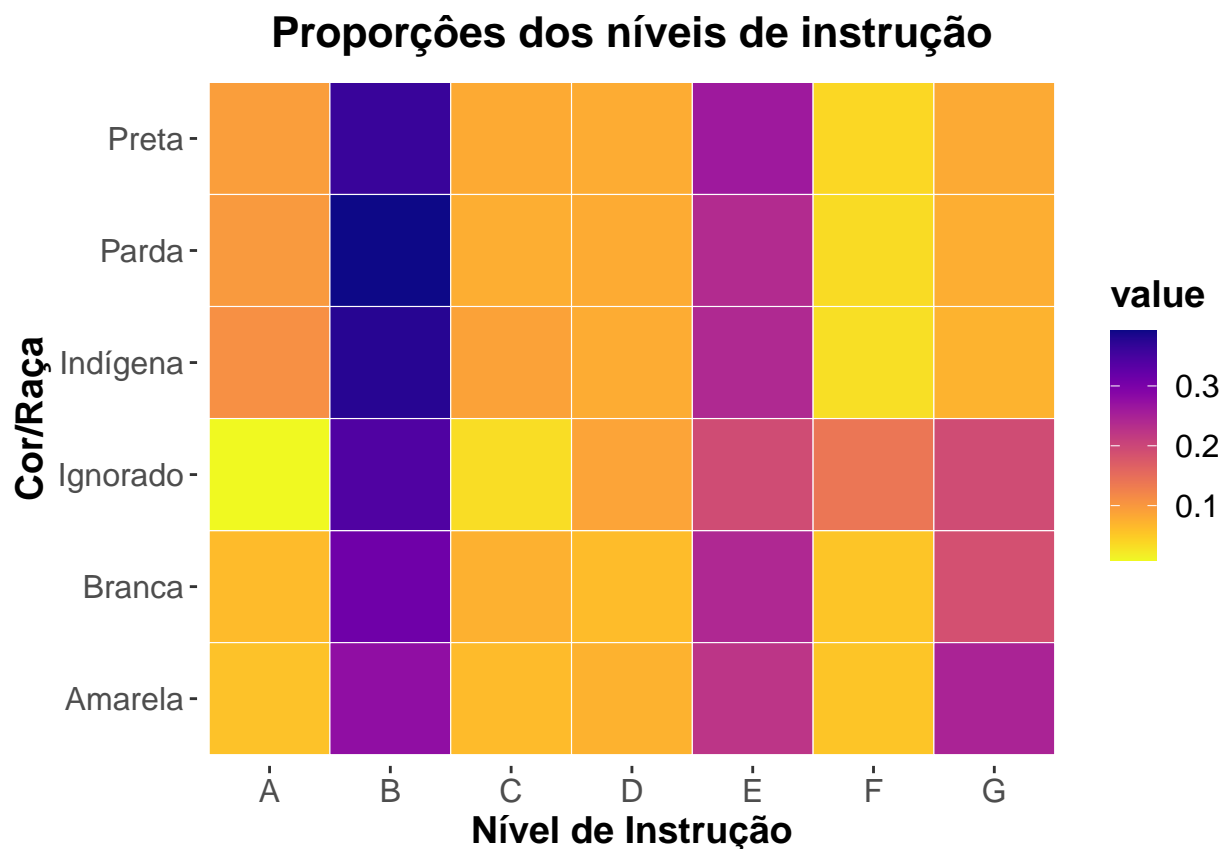
G - Superior completo

Na tabela abaixo, estudamos essa variável em subpopulações por sexo e Cor/Etnia a fim de observar possíveis contrastes nas proporções. Observa-se que, para níveis de instrução menores (A,B,C,D) há um percentual maior de homens do que mulheres, enquanto para níveis de instrução (E,F,G) há um percentual maior de mulheres.

Characteristic	Homem, N = 100,158,109	95% CI	Mulher, N = 109,431,498	95% CI
Nível de Instrução				
Sem instrução	8,003,192 (8.5%)	8.3%, 8.8%	8,355,626 (8.1%)	7.9%, 8.4%
Fundamental incompleto ou equivalente	34,831,677 (37%)	37%, 38%	34,644,920 (34%)	33%, 34%
Fundamental completo ou equivalente	7,470,408 (8.0%)	7.7%, 8.2%	7,714,804 (7.5%)	7.3%, 7.7%
Médio incompleto ou equivalente	7,295,016 (7.8%)	7.5%, 8.0%	7,069,539 (6.9%)	6.6%, 7.1%
Médio completo ou equivalente	22,040,691 (23%)	23%, 24%	25,494,440 (25%)	24%, 25%
Superior incompleto ou equivalente	4,043,992 (4.3%)	4.1%, 4.5%	4,679,410 (4.5%)	4.4%, 4.7%
Superior completo	10,131,363 (11%)	10%, 11%	14,967,603 (15%)	14%, 15%
Faltante	6,341,771		6,505,156	

Avaliando a distribuição desses níveis de instrução por Cor/Etnia, observa-se:

Nos níveis (A,B,C,D,E) há um contraste sutil entre indivíduos autodeclarantes brancos, amarelos em relação a indivíduos pretos, pardos e indígenas. A discrepância maior parece estar nos níveis de educação superior, onde essa diferença aumenta.



Caracterização da percepção de saúde da população brasileira

Um dos questionários da pesquisa busca informações sobre como a população avalia sua saúde, as perguntas consideradas foram:

- I00102: tem algum plano de saúde médico particular, de empresa ou órgão público?
- J007: Algum médico já deu o diagnóstico de alguma doença crônica, física ou mental, ou doença de longa duração (de mais de 6 meses de duração)?
- N001: Em geral, como o(a) Sr(a) avalia a sua saúde
- N00101: Considerando saúde como um estado de bem-estar físico e mental, e não somente a ausência de doenças, como você avalia o seu estado de saúde?

A tabela abaixo sumariza os resultados. É possível notar que grande parte da população brasileira não possui um plano de saúde, e mais de 30% parece ter sido diagnosticada com alguma doença. A percepção da saúde pelos entrevistados parece otimista para as duas últimas questões, com a grande maioria das respostas na faixa Regular - Muito Bom.

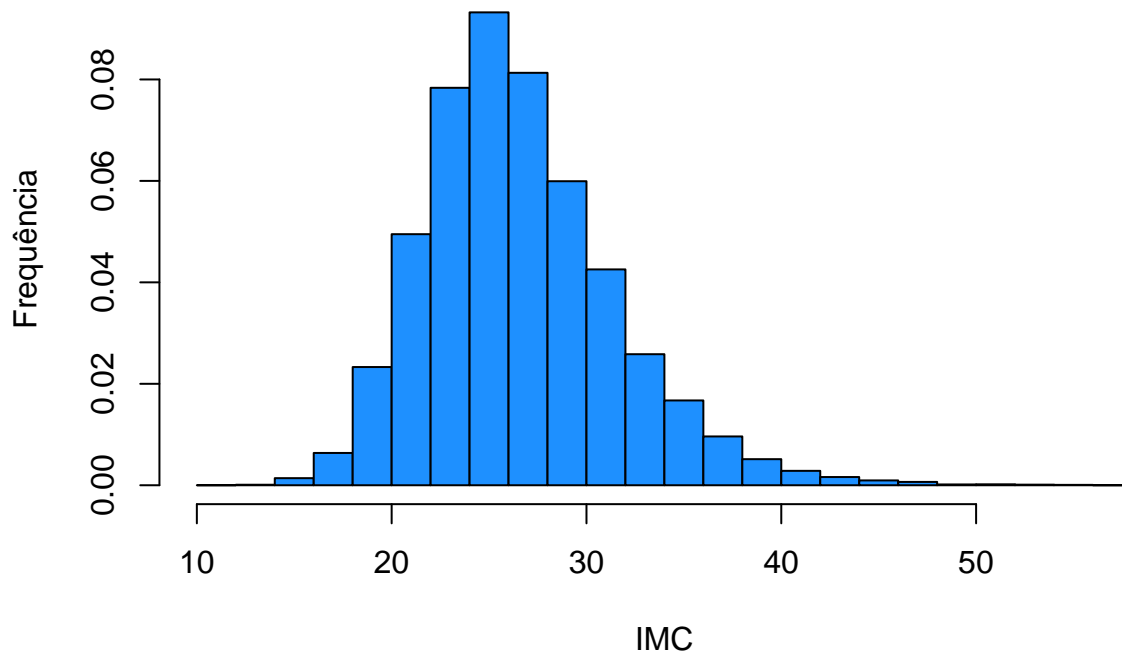
Characteristic	N = 209,589,607	95% CI
Possui um plano de saúde		
Sim	54,587,974 (26%)	25%, 27%
Não	155,001,633 (74%)	73%, 75%
Ignorado	0 (0%)	0.00%, 0.00%

Characteristic	N = 209,589,607	95% CI
Diagnóstico de Doença Crônica, Física ou Mental		
Sim	65,983,584 (31%)	31%, 32%
Não	143,606,023 (69%)	68%, 69%
Ignorado	0 (0%)	0.00%, 0.00%
Percepção geral da própria saúde		
Muito boa	10,734,669 (15%)	15%, 16%
Boa	35,394,282 (50%)	49%, 50%
Regular	20,448,416 (29%)	28%, 29%
Ruim	3,464,839 (4.9%)	4.7%, 5.1%
Muito ruim	907,252 (1.3%)	1.2%, 1.4%
Ignorado	0 (0%)	0.00%, 0.00%
Faltante	138,640,149	
Percepção da saúde considerando bem-estar físico/mental		
Muito boa	12,255,752 (17%)	17%, 18%
Boa	40,188,893 (57%)	56%, 57%
Regular	15,327,095 (22%)	21%, 22%
Ruim	2,576,685 (3.6%)	3.4%, 3.8%
Muito ruim	601,033 (0.8%)	0.76%, 0.95%
Ignorado	0 (0%)	0.00%, 0.00%
Faltante	138,640,149	

Caracterização dos hábitos de vida dos brasileiros.

Esta seção traz resultados do questionário contendo questões sobre os hábitos de alimentação e prática de exercícios físicos dos brasileiros. Na imagem abaixo vemos a distribuição dos IMC's na população-alvo.

IMC na População Estudada



Vemos que a média dos valores fica acima do valor 25 kg/m^2 , o que é um indicativo da prevalência de sobrepeso na população estudada.

Quanto aos hábitos de frequência semanal, os entrevistados foram interrogados com as seguintes questões:

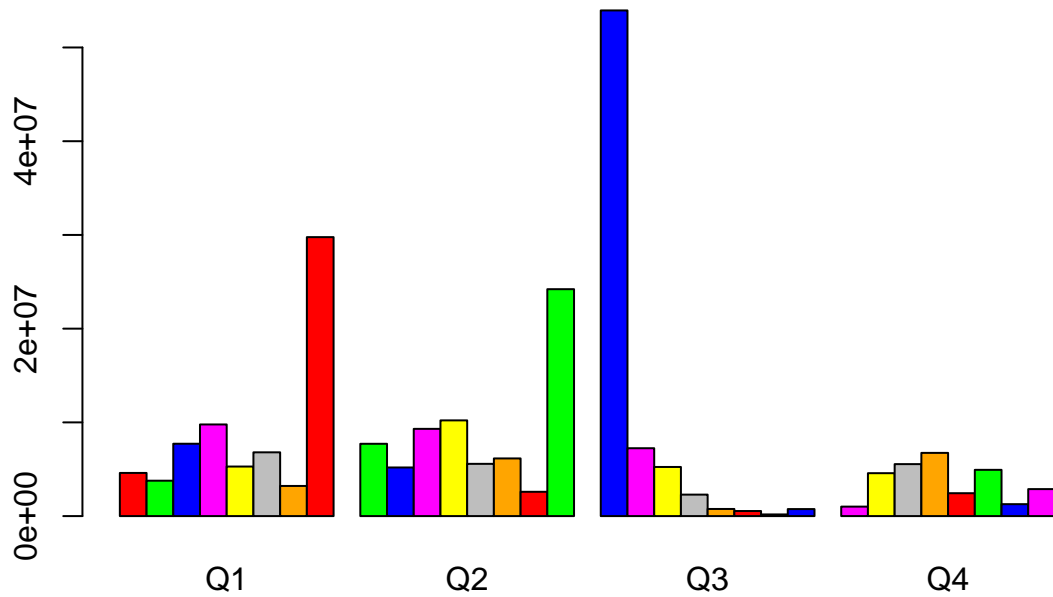
Q1: Em quantos dias da semana, o(a) Sr(a) costuma comer pelo menos um tipo de verdura ou legume (sem contar batata, mandioca, cará ou inhame) como alface, tomate, couve, cenoura, chuchu, berinjela, abobrinha?

Q2: Em quantos dias da semana o(a) Sr(a) costuma comer frutas?

– Q3: Em quantos dias da semana o(a) Sr(a) costuma substituir a refeição do almoço por lanches rápidos como sanduíches, salgados, pizza, cachorro quente, etc?

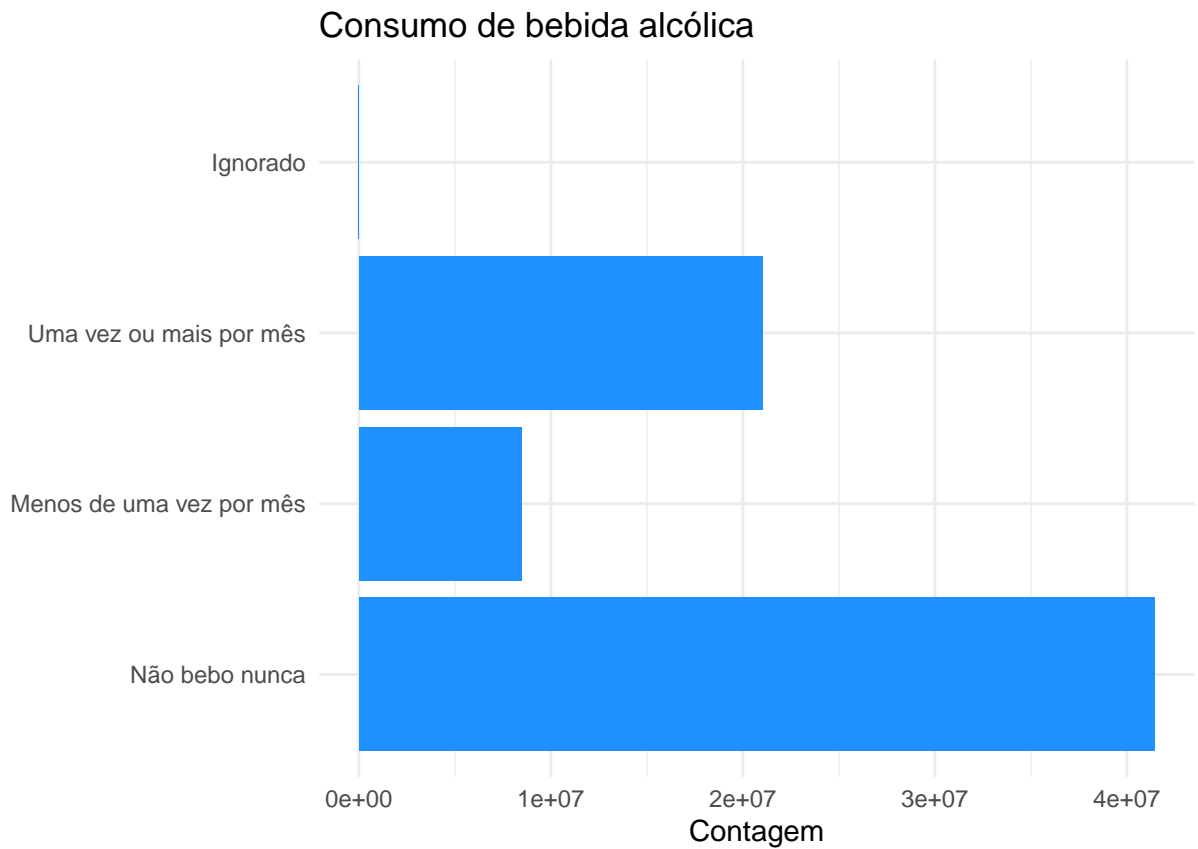
– Q4: Quantos dias por semana o(a) Sr(a) costuma (costumava) praticar exercício físico ou esporte?

Hábitos de frequência semanal

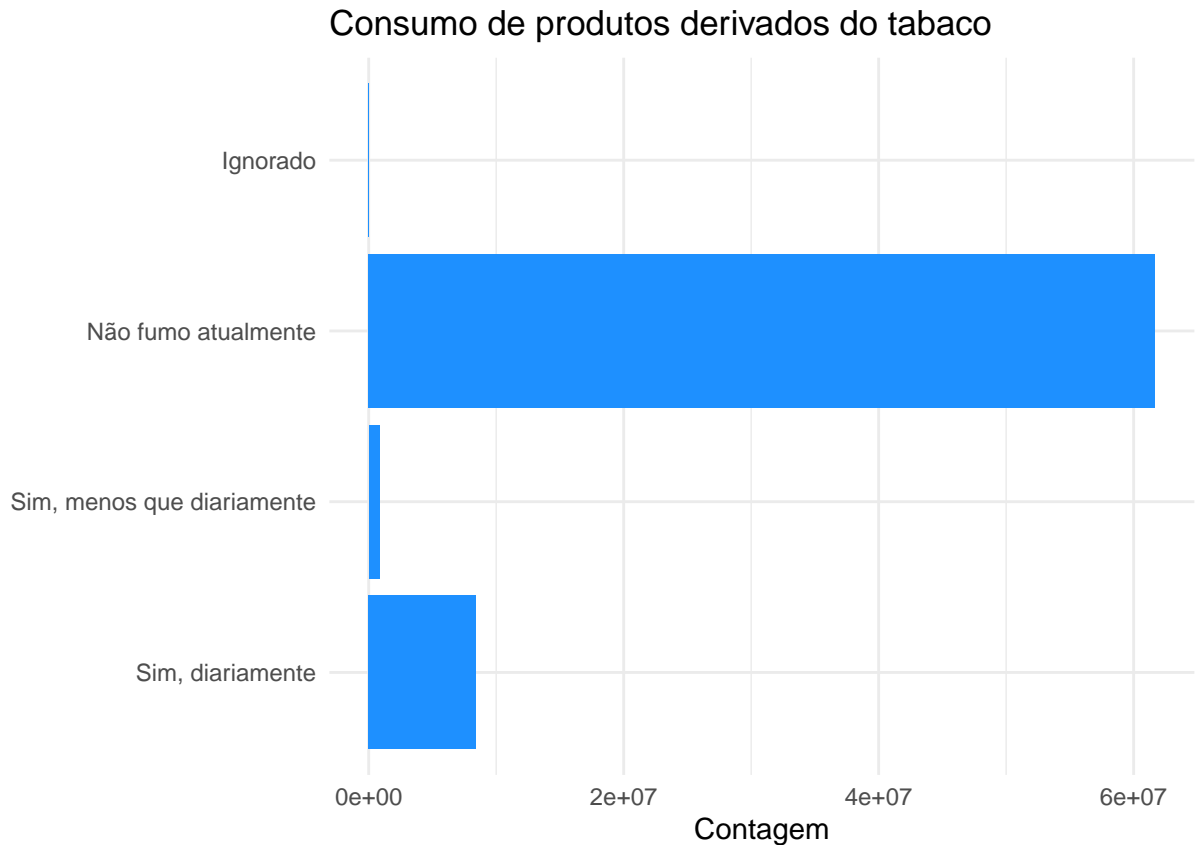


As três primeiras questões mostram uma frequência maior de respostas relacionadas a bons hábitos, como comer frutas/verduras 7 dias por semana e não substituir a alimentação diária por lanches rápidos. A quarta questão não mostra nenhum contraste marcante entre as respostas e obteve um número pequeno de contagens para todas as opções.

Analisando o padrão de consumo de bebidas alcólicas segundo as respostas, na imagem abaixo, vemos que há prevalência de respostas negando o consumo de álcool, embora as respostas afirmativas somem mais de 50% das respostas negativas.



Quanto ao consumo de produtos derivados do tabaco, vemos que há um número bem maior de respostas negativas em relação às respostas afirmativas, havendo prevalência de indivíduos que dizem não consumir.

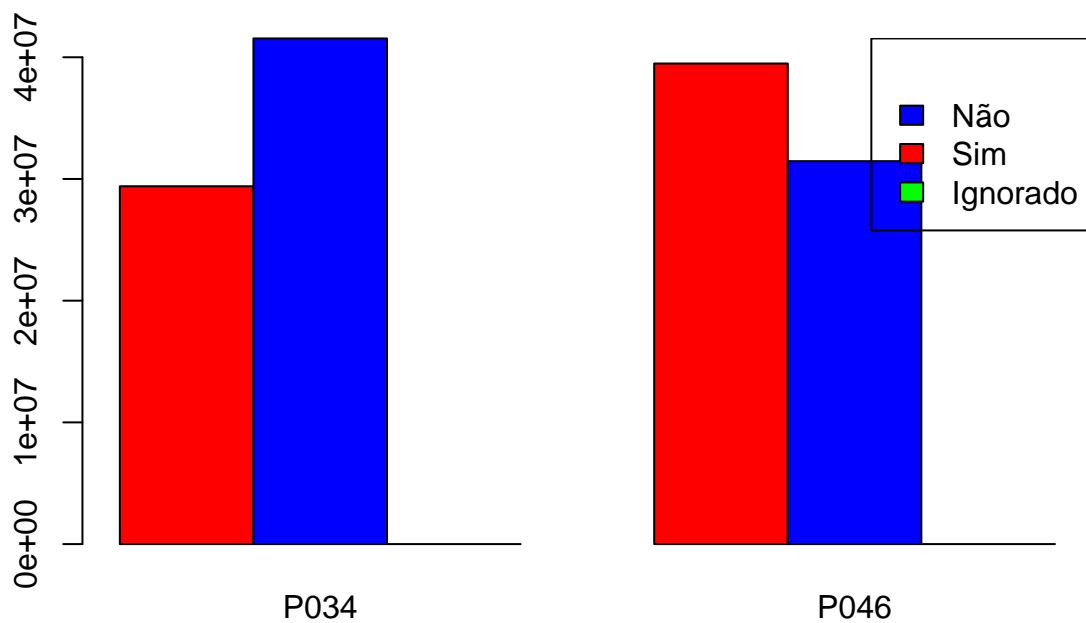


Avaliando o perfil de respostas para questões relacionadas à prática de exercícios:

P046: Perto do seu domicílio, existe algum lugar público (praça, parque, rua fechada, praia) para fazer caminhada, realizar exercício ou praticar esporte?

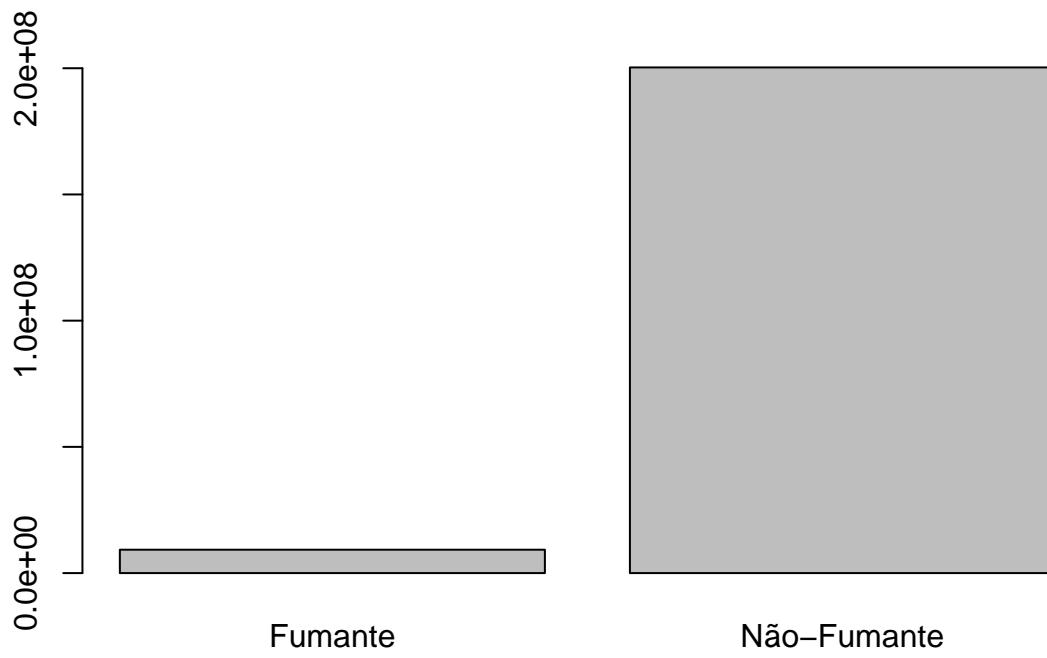
P034: Nos últimos três meses, o(a) Sr(a) praticou algum tipo de exercício físico ou esporte?

Vemos, pela imagem abaixo, que a contagem mostra que a grande maioria dos entrevistados não praticou exercícios físicos nos últimos meses, embora a maioria afirme ter um lugar próximo para praticar.

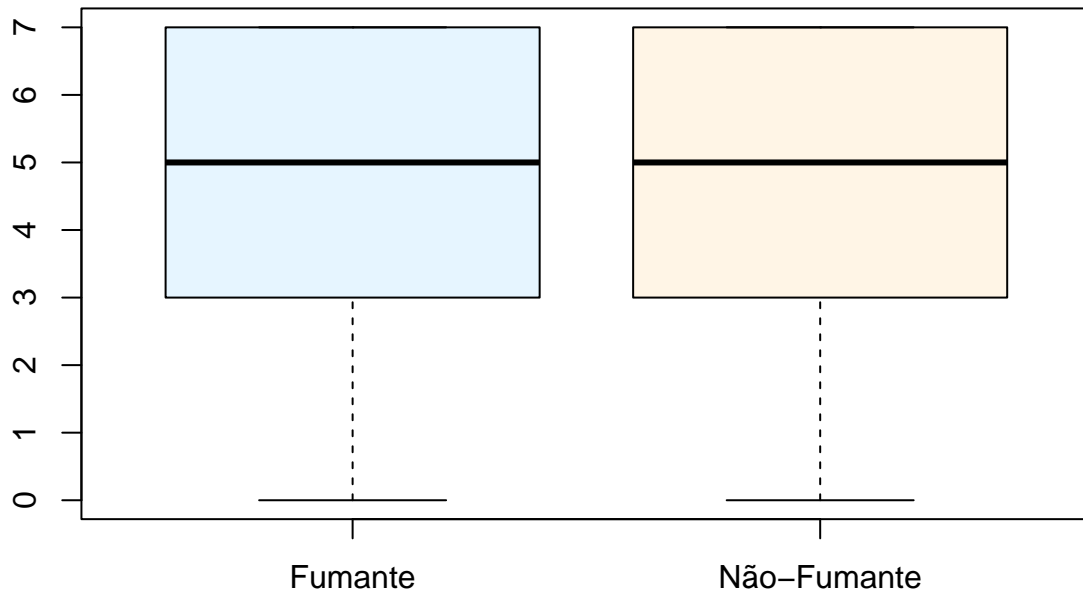


```
##
## Design-based one-sample t-test
##
## data: cbind(P046, P034) ~ 1
## t = 300.21, df = 7453, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  1.434034 1.579007
## sample estimates:
##      mean
## 1.44346
```

Com o t-teste acima, podemos comparar as duas variáveis e rejeitar a hipótese nula, ou seja, concluímos que não há uma relação significativa entre as variáveis. Sobre a prevalência do tabagismo no Brasil, vemos pela imagem abaixo que a maior parcela se declara não=fumante.



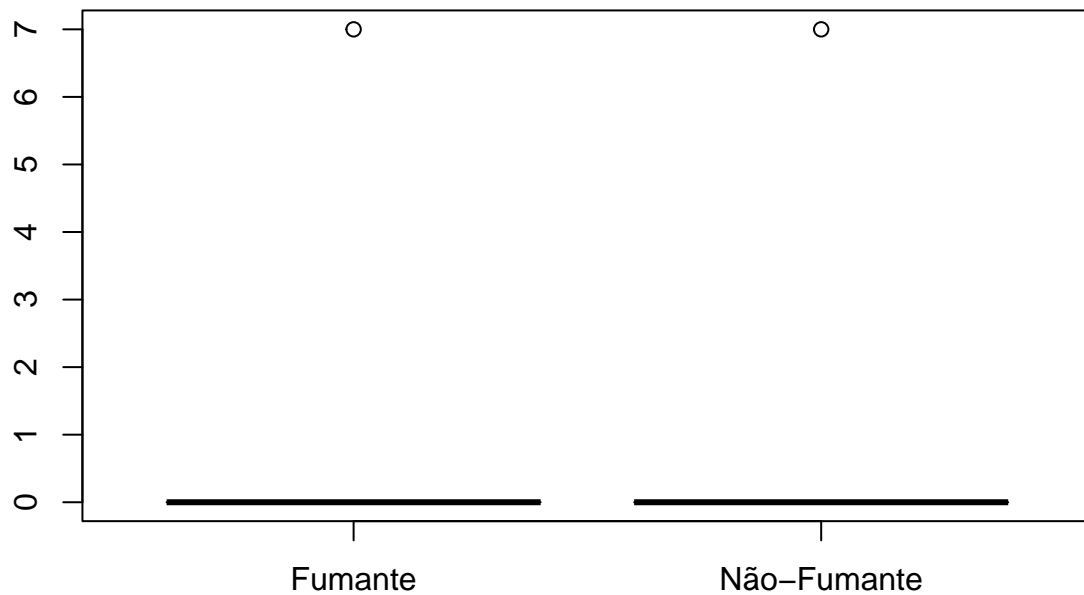
Alimentação saudável(Frequência semanal)



```
##
## Design-based t-test
##
## data: P00901 ~ fuma
## t = 9.2174, df = 7452, p-value < 2.2e-16
## alternative hypothesis: true difference in mean is not equal to 0
## 95 percent confidence interval:
##  0.2960785 0.4560310
## sample estimates:
## difference in mean
##      0.3760548
```

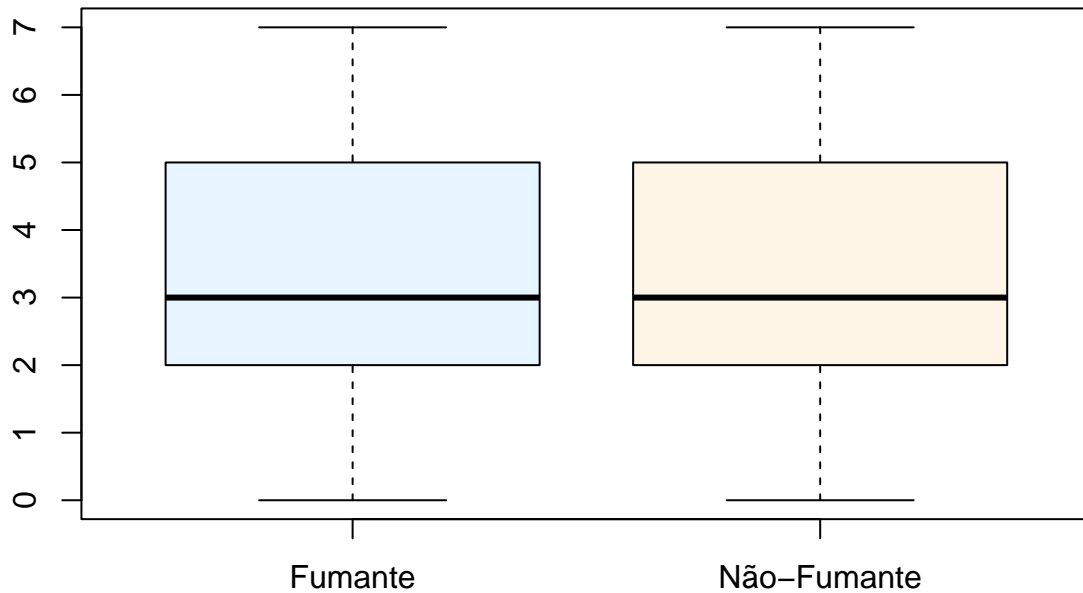
```
##
## Design-based t-test
##
## data: as.numeric(P018) ~ as.factor(fuma)
## t = 23.34, df = 7452, p-value < 2.2e-16
## alternative hypothesis: true difference in mean is not equal to 0
## 95 percent confidence interval:
##  0.973103 1.151545
## sample estimates:
## difference in mean
##      1.062324
```

Alimentação saudável(frequência semanal)



```
##  
## Design-based t-test  
##  
## data: P02602 ~ as.factor(fuma)  
## t = -6.5753, df = 7452, p-value = 5.183e-11  
## alternative hypothesis: true difference in mean is not equal to 0  
## 95 percent confidence interval:  
## -0.17664597 -0.09550894  
## sample estimates:  
## difference in mean  
## -0.1360775
```

Prática de exercício físico(frequência semanal)



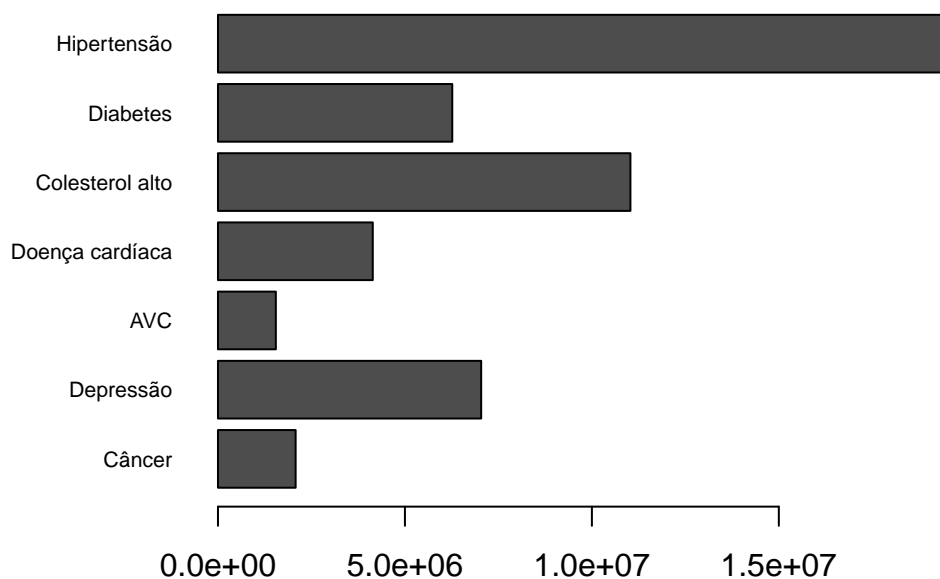
```
##
## Design-based t-test
##
## data: P035 ~ fuma
## t = 3.8772, df = 7270, p-value = 0.0001066
## alternative hypothesis: true difference in mean is not equal to 0
## 95 percent confidence interval:
##  0.1173468 0.3573460
## sample estimates:
## difference in mean
##           0.2373464
```

As análises boxplot acima mostram que a média para as respostas relacionadas a hábitos de alimentação e prática de exercício físico são as mesmas considerando as subpopulações de fumantes e não-fumantes. A conclusão dos t-testes é que não há relação significativa entre hábitos saudáveis e respostas afirmativas.

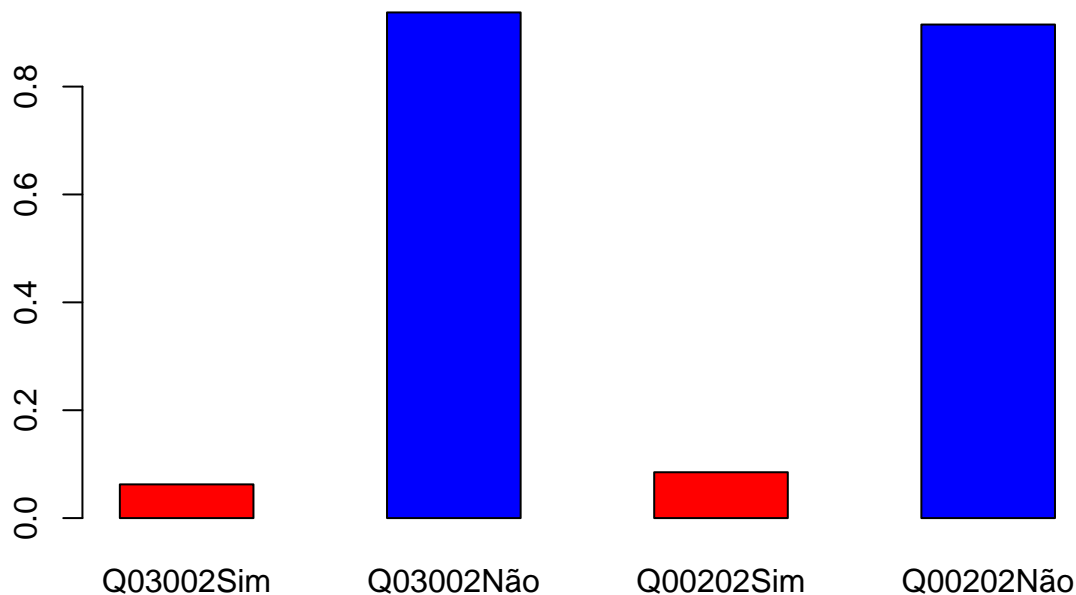
Caracterização da situação brasileira com relação às principais doenças crônicas

Ao perguntar aos entrevistados sobre diagnósticos médicos para doenças crônicas pode-se observar que há uma grande proporção de indivíduos que já receberam diagnóstico para ao menos uma doença crônica. A imagem abaixo mostra a frequência da incidência de cada doença considerada.

Characteristic	N = 209,589,607	95% CI
Doença crônica		
Não	249,390 (0.8%)	0.67%, 0.93%
Sim	31,178,843 (99%)	99%, 99%
Faltante	178,161,373	



Vemos que a maior ocorrência se dá para hipertensão e colesterol, doenças que podem ser relacionadas a maus hábitos alimentares. Quando perguntado se a ocorrência de Diabetes, Hipertensão tinha se dado no período gestacional (Questão Q00202), a proporção maior foi de ocorrências fora do período.



```
##           mean      SE
## publicaNão 0.013822 0.0052
## publicaSim 0.986178 0.0052

## medicacao
##      Sim
## 23960236
```

Vemos também pelas tabela acima que dentre os indivíduos com ocorrência de doenças crônicas, 100% diz ter recebido receita médica para um medicamento apropriado e há um grande percentual de indivíduos que adquirem esses medicamentos por serviços públicos.

CONCLUSÕES

A percepção geral da saúde da média dos brasileiros é positiva, embora a grande maioria não possua um plano de saúde. Os resultados do questionário sobre hábitos saudáveis mostra que boa parte dos entrevistados dizem manter boa alimentação. A prática de exercícios físicos parece não estar relacionada com a disponibilidade de locais públicos adequados. A média do IMC na população mostra sinais de ocorrência de obesidade. Também observou-se que o consumo de produtos do tabaco não tem relação com outros hábitos saudáveis. É possível que o efeito do entrevistador tenha gerado respostas não confiáveis sobre os hábitos dos entrevistados. Concluímos que há prevalência da ocorrência de doenças crônicas no Brasil, especialmente Hipertensão e Colesterol alto, o que pode ter relação com a distribuição do IMC na população e, portanto, obesidade. Uma pequena parcela dos afetados adquire essas doenças no período gestacional e a grande maioria utiliza medicamentos advindos de serviços públicos de saúde. A ampliação desses serviços, assim como construção

de mais áreas de lazer são parte dos esforços que devem compor as políticas que visem a ampliação do acesso a saúde e bem-estar no futuro dos brasileiros.