Análise, PNS - DIABETES

Data Prep

Diego Aguirre, Gabriel Rodrigues, Henrique Alves

2019

# Introdução

## R Markdown

## Leitura e preparação dos dados

Os dados da Pesquisa Nacional de Saúde 2013 são disponibilizados pelo IBGE no seu site.

**É importante lembrar que os dados relacionados a diabetes consideram pessoas com 18 anos ou mais**; por isso, o total de pessoas não é exatamente a população brasileira, isto é, perto de 200 milhões. mas é aproximadamente 143 milhões, com os dados de 2013.

### Variáveis interessantes para análise

Variáveis Demográficas e Pessoais:

* c006 - Sexo:
  + 1 - masculino
  + 2 - feminino
* c009 - Cor ou Raça
  + 1 - Branca
  + 2 - Preta
  + 3 - Amarela
  + 4 - Parda
  + 5 - Indígena
  + 9 - Ignorado
* d009 - curso mais avançado que cursou?
  + 01 - Classe de alfabetização – CA
  + 02 - Alfabetização de jovens e adultos
  + 03 - Antigo primário (elementar)
  + 04 - Antigo ginásio (médio 1º ciclo)
  + 05 - Regular do ensino fundamental ou do 1º grau
  + 06 - Educação de jovens e adultos (EJA) ou supletivo do ensino fundamental
  + 07 - Antigo científico, clássico etc. (médio 2º ciclo)
  + 08 - Regular do ensino médio ou do 2º grau
  + 09 - Educação de jovens e adultos (EJA) ou supletivo do ensino médio
  + 10 - Superior - graduação
  + 11 - Mestrado
  + 12 - Doutorado
  + - Não aplicável
* w00103 - peso
* w00203 - altura

Variáveis de Estilo de vida:

* p009 - quantos dias na semana o indivíduo come vegetais
* p018 - quantos dias na semana o indivíduo come frutas
* p020 - quantos dias na semana o indivíduo consome regrigerantes e sucos industriais
* p025 - quantos dias na semana o indivíduo come alimentos doces(bolos, tortas, chocolates…)
* p026 - quantos dias na semana o indivíduo substitui almoço ou janta por sanduiches, salgados ou pizzas.
* p028 - quantos dias na semana o indivíduo consome bebidas alcoolicas
* p035 - quantos dias por semana costuma praticar exerc físico ou esporte
* w00103 - peso final (medido) (kg)
* w00203 - altura final (medido) (cm)

Variáveis de Doenças Crônicas

* Q002 - Algum médico já lhe deu o diagnóstico de hipertensão arterial (pressão alta)?
  + 1 - Sim
  + 2 - Apenas durante a gravidez
  + 3 - Não Não aplicavel
* Q029 -
  + 1 - Há menos de 6 meses
  + 2 - Entre 6 meses e menos de 1 ano
  + 3 - Entre 1 ano e menos de 2 anos
  + 4 - Entre 2 anos e menos de 3 anos
  + 5 - Há 3 anos ou mais
  + 6 - Nunca fez Não aplicavel
* Q030 - Algum médico já lhe deu o diagnóstico de diabetes? (CLASSE A SER PREDITA)
  + 1 - Sim
  + 2 - Apenas durante a gravidez (só para mulheres)
  + 3 - Não Não aplicavel
* Q060 - Algum médico já lhe deu o diagnóstico de colesterol alto?
  + 1 - Sim
  + 2 - Não Não aplicavel
* Q068 - Algum médico já lhe deu o diagnóstico de AVC (Acidente Vascular cerebral) ou derrame?
  + 1 - Sim
  + 2 - Não Não aplicavel
* Q124 - Algum médico já lhe deu o diagnóstico de insuficiência renal crônica?
  + 1 - Sim
  + 2 - Não Não aplicavel
* R039 -R39. Durante a sua vida, a sra já ficou grávida (mesmo que a gravidez não tenha chegado até o final)?
  + 1 - Sim
  + 2 - Não Não aplicavel

Variáveis de Identificação e controle

V0001 - Unidade da Federação V0026 - Tipo de situação censitária

#Utilizando o pacote `srvyr`  
library(srvyr)  
library(ggplot2)  
library(ggpubr)  
library(dplyr)  
library(knitr)  
library(tidyr)

## Criação de variáveis adicionais

### IMC – Índice de Massa Corporal

Ao invés de utilizarmos diretamente o peso como uma variável de comparação, vamos utilizar o Índice de Massa Corporal (IMC), definido como:

Vamos considerar também as faixas do IMC definidas pela Organização Mundial da Saúde, conforme mostrado na tabela a seguir:

|  |  |  |
| --- | --- | --- |
| Classificação | Faixa de Peso | Sintomas |
| Muito abaixo do peso | 16 a 16,9 kg/m2 | Queda de cabelo, infertilidade, ausência menstrual |
| Abaixo do peso | 17 a 18,4 kg/m2 | Fadiga, stress, ansiedade |
| Peso normal | 18,5 a 24,9 kg/m2 | Menor risco de doenças cardíacas e vasculares |
| Acima do peso | 25 a 29,9 kg/m2 | Fadiga, má circulação, varizes |
| Obesidade Grau I | 30 a 34,9 kg/m2 | Diabetes, angina, infarto, aterosclerose |
| Obesidade Grau II | 35 a 40 kg/m2 | Apneia do sono, falta de ar |
| Obesidade Grau III | maior que 40 kg/m2 | Refluxo, dificuldade para se mover, escaras, diabetes, infarto, AVC |

### Faixa etária

Uma outra classificação interessante é por faixa etária. Temos uma classificação publicada pela Revista Exame no artigo de Novembro de 2016 baseado nos dados da PNAD 2015 [Um retrato do Brasil e do brasileiro, segundo o IBGE](https://exame.abril.com.br/brasil/um-retrato-do-brasil-e-do-brasileiro-segundo-o-ibge/)

No artigo mencionado são considerados os seguintes grupos etários:

|  |  |
| --- | --- |
| Faixa | Classe Aqui |
| 0–4 | 2 |
| 5–9 | 7 |
| 10–14 | 12 |
| 15–19 | 18 |
| 20–24 | 22 |
| 25–39 | 32 |
| 40–59 | 50 |
| 60+ | 60 |

## Tabulações

### Prevalência de pessoas com diabetes

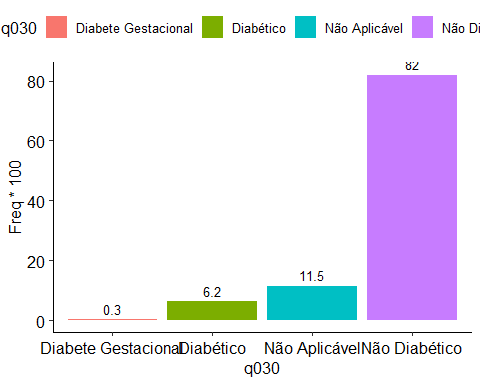
Para verificarmos se nossas métricas estão corretas buscamos referência sobre a análise de diabetes pela PNS através de outras fontes:

**No Brasil, conforme narra o artigo, dados recentes da Pesquisa Nacional em Saúde (PNS) de 2013 estimou a prevalência de DM em 6,2% dos participantes com 18 anos ou mais.**

Fiocruz: <https://portal.fiocruz.br/noticia/diabetes-pesquisa-avalia-os-fatores-associados-qualidade-de-vida>

Esse valor demonstra a consistência das nossas análises como vemos no gráfico abaixo onde a taxa de diabéticos correspondeu ao do estudos publicados pela Fiocruz.

Utilizamos a variável q030 que identifica se o participante já foi diagnosticado com Diabetes por algum médico. Os valores não aplicável demonstra as pessoas que nunca fizeram exames de sangue, definidas pela variável q029 que pergunta quando foi a última vez que a pessoa fez exame de sangue.

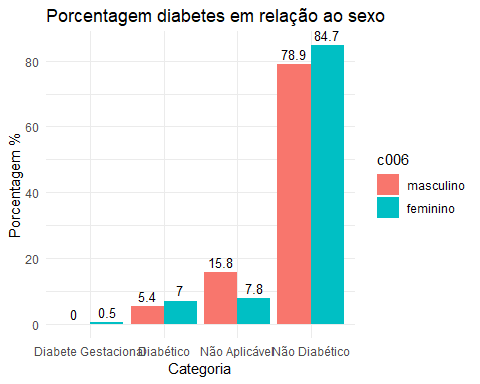


### Porcentagem de Diagnóstico de Diabetes por sexo

Em termos de distribuição da taxa de diabéticos em relação ao sexo chegamos a uma taxa de 5,4% para os homens e 7% para as mulheres. Ambos os resultados correspondem com a análise do artigo abaixo.

Para o cálculo utilizamos a variável c006 que representa o sexo do entrevistado e a variável q030 utilizada na análise anterior.

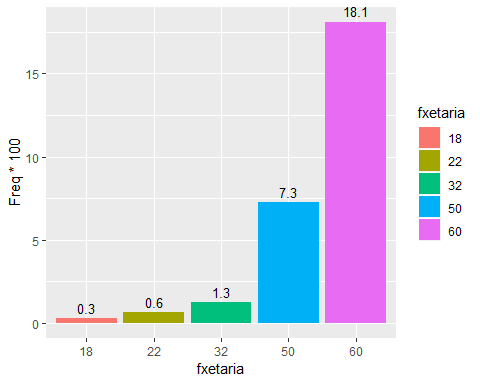
Artigo: <http://www.scielo.br/scielo.php?pid=s2237-96222015000200305&script=sci_abstract&tlng=pt>



### Procentagem de diabéticos por Faixa Etária

No atlas do IDF de 2017 mostrou-se que na população mundial acima dos 65 anos a taxa de prevalência de diabetes é de 19%, bem próximo da nossa análise de 18,1% para a população com 60 anos ou mais.

fonte: <https://www.diabetes.org.br/profissionais/images/2018/poster-atlas-idf-2017.pdf>



### Frequência de Atividade Física

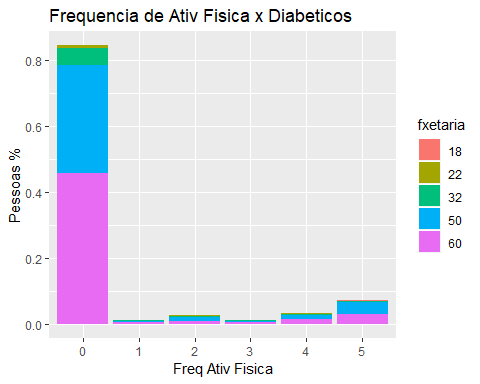
Para permitir uma melhor análise da relação da atividade física com a diabetes, criamos uma nova variável que permita relacionar a frequência de modo direto, já que a variável original m016 apresenta uma codificação não invertida.

|  |  |  |  |
| --- | --- | --- | --- |
| Valor Original | Codificação Original | Novo Valor | Codificação Variável freqativ |
| 1 | mais de uma vez por semana | 5 | mais de uma vez por semana |
| 2 | uma vez por semana | 4 | uma vez por semana |
| 3 | de 2 a 3 vezes por mês | 3 | de 2 a 3 vezes por mês |
| 4 | algumas vezes no ano | 2 | algumas vezes no ano |
| 5 | uma vez no ano | 1 | uma vez no ano |
| 6 | nenhuma vez | 0 | nenhuma vez |

### Taxa de diabéticos por frequencia de atividade física

Um dos fatores preditores que podem indicar se uma pessoa é propensa ou não a desenvolver a diabetes é se ela pratica ou não atividade física. Mais de 80% dos diabéticos não praticam atividades físicas, mas vemos que há um aumento dos diabéticos que mantém uma frequencia de 1 vez por semana ou mais. Como demonstrado no artigo abaixo, realizar atividades físicas é essencial para ajudar no bom controle da doença e garantir uma boa qualidade de vida.

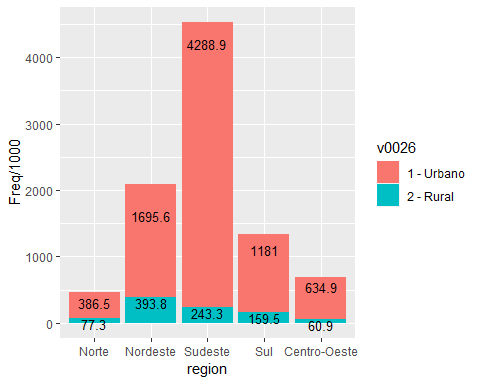
Artigo: <https://www.diabetes.org.br/publico/meu-esporte-minha-vida/965-atividade-fisica-e-diabetes>



## Taxa de diabéticos por região e por situação censitária

## Regiao Freq\_Urbano Freq\_Rural  
## 1 Norte 8519136 2366831.9  
## 2 Nordeste 29696279 9251295.4  
## 3 Sudeste 59677481 4397201.2  
## 4 Sul 18410830 3213833.5  
## 5 Centro-Oeste 9828695 946873.7

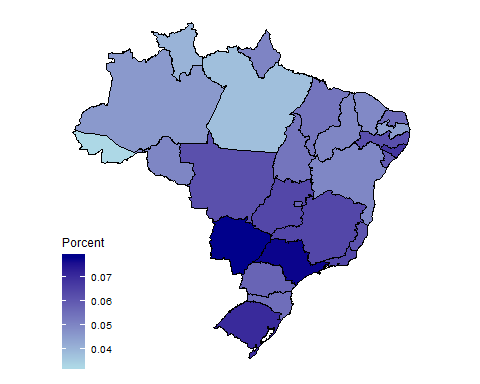
Dos aproximadamente 146 milhões de brasileiros acima de 18 anos, cerca de 64 milhões(44%) estão no sudeste, isso implica que a quantidade de diabéticos nessa região vai muito maior do que as outras regiões como vemos no gráfico abaixo. Cerca de 49.7% dos diabéticos residem no sudeste brasileiro.



## Distribuição das taxas de diabéticos por estado

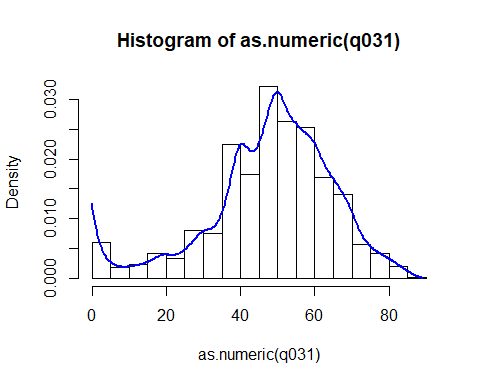
Como demonstrado no artigo acima as maiores taxas de diabéticos do Brasil estão em Mato Grosso do Sul, São Paulo e Rio Grande do Sul respectivamente. Esses dados correspondem com as taxas demonstrados no gráfico abaixo.

## Warning: package 'brazilmaps' was built under R version 3.5.3



## Distribuição da idade de diagnóstico de Diabetes

Temos o gráfico que demonstra a idade que as pessoas foram diagnosticadas com a doença. Acima dos 30 anos é que os diagnósticos são mais realizados e tendo seu pico próximo aos 50 anos. Como idade é um dos fatores preditivos, faz sentido que com o passar dos anos a quantidade de diagnósticos aumente.



## Data Prep

### Preenchendo NAs

* Variáveis d009, p028, p035 e r039 que eram NAs serão preenchidas com 0, já que demonstram uma ausência de seus valores.
* Variável c006(peso), será preenchida com a média das pessoas separadas pelo sexo(c006) e pela faixa etária(fxetaria).
* Variável w00203(altura), será preenchida com a média das pessoas separadas pelo sexo(c006) e pela faixa etária(fxetaria).

### Inversão de valores de variáveis

Para permitir e simplificar a análise a partir de um modelo alteraremos os valor de variáveis.

q002, pressão alta:

|  |  |  |  |
| --- | --- | --- | --- |
| Valor Original | Codificação Original | Novo Valor | Codificação Variável pressao\_alta\_q002 |
| 1 | sim | 1 | sim |
| 2 | apenas na gravidez | 1 | sim |
| 3 | não | 0 | não |

q060, colesterol alto:

|  |  |  |  |
| --- | --- | --- | --- |
| Valor Original | Codificação Original | Novo Valor | Codificação Variável colesterol\_alto\_q060 |
| 1 | sim | 1 | sim |
| 2 | não | 0 | não |

q124, insuficiencia renal:

|  |  |  |  |
| --- | --- | --- | --- |
| Valor Original | Codificação Original | Novo Valor | Codificação Variável insuficiencia\_renal\_q124 |
| 1 | sim | 1 | sim |
| 2 | não | 0 | não |

p050, fumante: definição de fumante: <http://pepsic.bvsalud.org/scielo.php?script=sci_arttext&pid=S1413-03942014000200015>

|  |  |  |  |
| --- | --- | --- | --- |
| Valor Original | Codificação Original | Novo Valor | Codificação Variável fumante\_p050 |
| 1 | sim, diariamente | 1 | sim |
| 2 | sim, menos que diariamente | 0 | não |
| 3 | não fumo atualmente | 0 | não |

q068, avc:

|  |  |  |  |
| --- | --- | --- | --- |
| Valor Original | Codificação Original | Novo Valor | Codificação Variável ja\_ficou\_gravida\_r039 |
| 1 | sim | 1 | sim |
| 2 | não | 0 | não |

q068, gravidez:

|  |  |  |  |
| --- | --- | --- | --- |
| Valor Original | Codificação Original | Novo Valor | Codificação Variável ja\_ficou\_gravida\_r039 |
| 1 | sim | 1 | sim |
| 2 | não | 0 | não |

# Decisão

Definir remover os registros que tem como NA diagnóstico de colesterol alto q060, pressão alta q002 e diabetes q030.

## [1] "quantidade total de diabéticos: 9121631.31704213"

## [1] "perda de diabéticos: 177395.69150922"

## [1] "porcentagem de perda de diabeticos: 0.0194478032868734"

## [1] ""

## [1] "quantidade total : 146308458.121465"

## [1] "perda total: 24344038.5927866"

## [1] "porcentagem de perda total: 0.166388456999364"