

**Curso Intermediário (apoiado por *software* R) da Análise da Situação de Saúde aplicado a Emergências Sanitárias, com foco na COVID-19**

**Aula 9 - Calculando taxas de incidência**

## **Apresentação**

Parabéns! Você chegou à última aula do nosso curso. Ao longo das aulas anteriores, você aprendeu conceitos de epidemiologia e bioestatística descritivas. Esses conceitos serão importantes para que você comece a fazer análises de saúde utilizando o R.

Para a última aula do módulo, vamos aprender a calcular taxas de incidência. Essas medidas são úteis quando queremos comparar períodos ou territórios diferentes em análise da situação de saúde. Essas medidas são primordiais na condução e na avaliação de surtos, epidemias e pandemias, como é o caso da situação ocorrida desde 2019, com a Covid-19.

Em nossa última aula, vamos aprender a relacionar dados de fontes diferentes no R. Vamos trazer informações sobre população e sobre os casos, para que possamos calcular a incidência da doença em diferentes estados considerando os dados fornecidos.

Para essa aula, estamos fornecendo um recorte de 299 casos de dados do e-SUS Notifica do Distrito Federal e um recorte de 299 casos de dados do Sivep-Gripe. Esses dados foram utilizados em todas as aulas anteriores e serão utilizados mais uma vez para criarmos indicadores de saúde.

Agradecemos por ter chegado até aqui! Isso mostra seu esforço e dedicação a esse curso. Parabenizamos você e torcemos para que tenha aprendido o máximo possível com esse curso.

## **1. Introdução**

Quando falamos de taxa de incidência, estamos falando de um indicador de saúde. Escolhemos a taxa de incidência pela facilidade de calcular esse indicador com os recursos utilizados nesse curso. Seu cálculo pode ser reproduzido para qualquer taxa que for utilizada na área da saúde.

Uma parte importante dos indicadores de saúde é compreender como os eventos de saúde se comportam no espaço e no tempo. Isso significa que é necessário entender se as medidas de morbidade e mortalidade estão se comportando de forma a aumentar ou diminuir o fenômeno em estudo (1).

Mas como podemos comparar as medidas de saúde? Em aulas anteriores aprendemos que não é possível comparar medidas absolutas. Quando pensamos na ocorrência de eventos de saúde nas populações dos estados, por exemplo, é de se esperar que o número de casos de uma doença em São Paulo seja maior que o número de casos da mesma doença em Alagoas. Por que isso acontece? Porque a população de São Paulo é mais numerosa que a população de Alagoas.

Aprendemos que para comparar os eventos precisamos trabalhar com frequências relativas, ou seja, frequências que representam uma fração proporcional a um conjunto de dados. É possível comparar medidas de tendência central e medidas de dispersão de diferentes populações, mas quando fazemos a avaliação da frequência de doenças, essas medidas podem não estar disponíveis, sendo necessário trabalhar com outras medidas (2).

Nesse sentido, quando queremos comparar diferentes espaços e diferentes períodos, devemos usar taxas ou coeficientes, que aqui serão trabalhados como sinônimos. Quando falamos de taxas em epidemiologia, estamos nos referindo ao número de ocorrências de determinado agravo em relação à população considerada para aquele território e período.

Por exemplo, como sabemos se a incidência de dengue aumentou dos anos 1990 para 2020 no Brasil? O número pode ter aumentado de forma proporcional à população, tornando inviável a análise somente pelo número de casos. Nesse caso, calculamos a taxa de incidência, ou seja, o número de casos por 100.000 habitantes em cada ano e comparamos para verificar se houve aumento ou decréscimo.

## 2. Principais taxas na epidemiologia

No campo da epidemiologia, temos dois principais grupos de indicadores: os indicadores de morbidade e os indicadores de mortalidade (1). Esses indicadores nos ajudam a compreender o comportamento de doenças e evoluções em saúde.

No campo da morbidade, encontramos duas grandes medidas epidemiológicas: a prevalência e a incidência. A prevalência se preocupa em mensurar todos os casos de uma determinada doença, novos e antigos, em um recorte definido de tempo e espaço. A incidência também ocorre para delimitar um recorte de período e de tempo, mas se preocupa em mensurar apenas casos novos. Podemos calcular taxas para as duas medidas:

- Taxa de incidência: mensura o número de casos novos de uma determinada doença, em determinado período e local, dividido pela população e ponderado por alguma potência de 10, em geral 100.000. Sua fórmula é dada por:

$$\begin{aligned} & \textit{Taxa de incidência} \\ &= \frac{\textit{número de casos novos em determinado período e local}}{\textit{população em determinado período e local}} \times 100.000 \end{aligned}$$

- Taxa de prevalência: mensura o número de casos atuais de uma determinada doença dividido pela população e ponderado por alguma potência de 10, em geral 100.000. Sua fórmula é dada por:

$$\begin{aligned} & \textit{Taxa de prevalência} \\ &= \frac{\textit{número de casos em determinado período e local}}{\textit{população em determinado período e local}} \times 100.000 \end{aligned}$$

No campo da mortalidade, temos mais de uma medida que ajuda a mensurar os óbitos e a letalidade de uma doença. Entretanto, nessa aula, vamos nos prender à taxa de mortalidade.

- Taxa de mortalidade: mensura o número de óbitos por alguma doença, dividido pela população e ponderado por uma potência de 10, em geral 1.000. Sua fórmula é dada por:

$$\begin{aligned} & \textit{Taxa de mortalidade} \\ &= \frac{\textit{número de óbitos em determinado período e local}}{\textit{população em determinado período e local}} \times 1.000 \end{aligned}$$

Os mecanismos para o cálculo de taxas no R é o mesmo. Nessa aula ensinaremos como calcular a taxa de incidência, mas os procedimentos podem ser reproduzidos para quaisquer outras taxas.

### 3. Carregando pacotes e importando os dados

Para essa aula, usaremos o pacote *Tidyverse* e as mesmas bases de dados utilizados nas aulas anteriores. Também vamos rodar novamente o *script* da aula 3, que realiza toda a padronização dos dados. Para isso usaremos os seguintes comandos:

```
# carregando pacotes
library(tidyverse) #tratamento dos dados

## datasets de dados
esus = read.csv2("dados/20210601_dadosesus_df.csv") #importando dados do esus
df de 01/06/2021
sivep = read.csv2("dados/20210823_dadossivep.csv") #importando dados do esus df
de 23/08/2021

source("scripts/03_aula_importando_dados.R") #realizando os tratamentos do
script 03
```

Os dados serão importados e tratados.

### 4. Importando dados populacionais

Além dos dados que trabalhamos nas aulas anteriores, vamos precisar dos dados populacionais. Esses dados podem ser obtidos pelo *TabNet*, a versão online do tabulador de dados do DATASUS, que pode ser usada para a criação de indicadores de saúde.

Estamos disponibilizando a população estimada mais recente, que foi baixada pelo TabNet, para cada UF do Brasil, no ano de 2019. Os dados foram cedidos pelo IBGE. Para importar esses dados, usaremos o seguinte comando.

```
# datasets de população
pop = read.csv2("dados/20211010_populacao.csv") #importando dados
populacionais do Tabnet
```

Assim como os dados de casos de Covid-19, os dados populacionais serão baixados. Nós temos três variáveis nessa tabela de dados: o código da UF, o nome da UF e o número de pessoas, conforme a imagem:

09_taxas_epidemiologicas.R x		pop x	
Filter			
	COD	UF	POP
1	11	Rondônia	1777225
2	12	Acre	881935
3	13	Amazonas	4144597
4	14	Roraima	605761
5	15	Pará	8602865
6	16	Amapá	845731
7	17	Tocantins	1572866
8	21	Maranhão	7075181

## 5. Calculando taxas de incidência

Nesse exercício, calcularemos as taxas de incidência com o total de notificações por UF, apenas a título de exercício, dado que o número de observações é pequeno. Nas rotinas de vigilância, você deverá considerar apenas os casos confirmados da doença na criação de taxas de incidência. Além disso, deverá fazer o recorte temporal, considerando o período com base na data de primeiros sintomas da doença.

Para calcular as taxas de incidência, precisaremos unir os dados do Sivep-Gripe aos dados populacionais. Para isso usaremos o comando *“merge”*, que faz parte das funções básicas do R. Esse comando permite que realizemos a união de duas tabelas diferentes por meio de uma coluna com a variável identificadora nos dois espaços.

Nesse caso, vamos usar o código do estado no IBGE, variável que já se encontra disponível na tabela de dados populacionais. Na tabela de dados do Sivep-Gripe, esse dado não está disponível, entretanto conseguimos obtê-lo a partir dos códigos de município disponíveis em outra variável. Por isso, criaremos uma variável com os dois primeiros dígitos dos códigos de município, para que possamos calcular o número de notificações por UF, conforme o comando:

```
# criando tabela de casos
```

```
## importante, ao calcular a taxa de incidência, usamos apenas casos confirmados.  
## por causa do n baixo, usaremos todas as notificações nesse exercício
```

```
sivep$COD = as.character(substring(sivep$CO_MUN_RES, 1, 2)) #incluindo código da UF
```

Após isso, temos uma nova coluna em nossa tabela, com os códigos de UF, conforme a imagem:

DT_RES	RES_JGG	RES_JGM	RES_JGA	ESTRANG	VACINA_COV	DOSE_1_COV	DOSE_2_COV	LAB_PR_COV	LOTE_1_COV	LOTE_2_COV	FNT_IN_COV	COD
	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	29
	4	4	4	NA	NA	NA	NA	NA	NA	NA	NA	27
	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	35
	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	35
	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	35
	NA	NA	NA	NA	2	NA	NA	NA	NA	NA	1	35
	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	52
	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	35
	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	29
12/01/2021	1	1	4	NA	2	NA	NA	NA	NA	NA	1	35

Após isso, vamos criar uma tabela de frequência com os dados de notificações por UF. Também daremos nome para as colunas dessa nova tabela e vamos transformar os códigos de UF em uma variável categórica, conforme os comandos:

```
casos = data.frame(table(sivep$COD))
colnames(casos) = c("COD", "casos")
casos$COD = as.character(casos$COD)
```

Assim, obteremos uma tabela com o número de casos por UF, sendo que as UF estão representadas por seus códigos no IBGE. A saída é a seguinte:

	COD	casos
1	11	1
2	12	1
3	13	12
4	15	4
5	17	2
6	22	2
7	23	14
8	24	2
9	25	9
10	26	5
11	27	4
12	28	5
13	29	11

Para relacionar essa tabela com a tabela de população, será necessário transformar o código do IBGE daquela tabela em uma variável categórica, assim como fizemos na tabela anterior.

```
pop$COD = as.character(pop$COD) #transformando o código da UF em character
```

Após essa transformação, faremos a união das duas tabelas, usando o comando *merge*:

```
# unindo os dados em um único data.frame  
taxa = merge(casos, pop, by = "COD")
```

Após isso, teremos uma tabela com o código do IBGE, o número de notificações, o nome da UF e a população da UF:



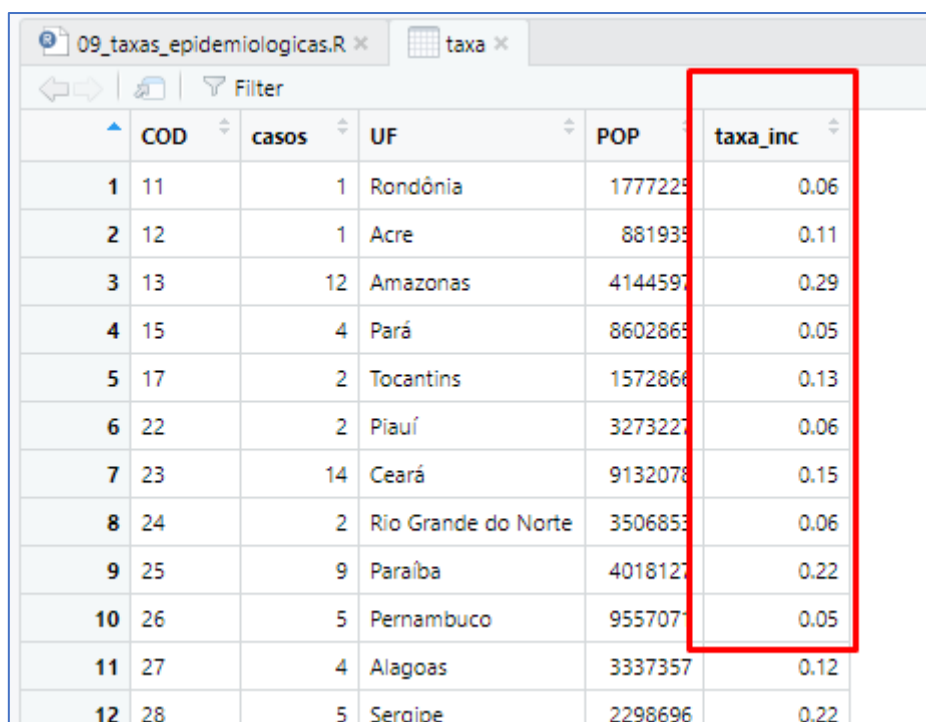
	COD	casos	UF	POP
1	11	1	Rondônia	1777225
2	12	1	Acre	881935
3	13	12	Amazonas	4144597
4	15	4	Pará	8602865
5	17	2	Tocantins	1572866
6	22	2	Piauí	3273227
7	23	14	Ceará	9132078
8	24	2	Rio Grande do Norte	3506853
9	25	9	Paraíba	4018127

Com esses dados, podemos calcular a taxa de incidência utilizando o comando a seguir:

```
#calculando taxa de incidência por UF  
taxa$taxa_inc = round(taxa$casos/taxa$POP*100000, 2)
```

Com isso, criaremos uma coluna em nossa tabela com a taxa de incidência calculada, que representa o número de notificações novas para cada 100.000 habitantes de cada estado, conforme a imagem:





09\_taxas\_epidemiologicas.R x taxa x

Filter

	COD	casos	UF	POP	taxa_inc
1	11	1	Rondônia	1777225	0.06
2	12	1	Acre	881935	0.11
3	13	12	Amazonas	4144597	0.29
4	15	4	Pará	8602865	0.05
5	17	2	Tocantins	1572866	0.13
6	22	2	Piauí	3273227	0.06
7	23	14	Ceará	9132078	0.15
8	24	2	Rio Grande do Norte	3506853	0.06
9	25	9	Paraíba	4018127	0.22
10	26	5	Pernambuco	9557071	0.05
11	27	4	Alagoas	3337357	0.12
12	28	5	Sergipe	2298696	0.22

## 6. Encerramento

Parabéns! Você finalizou o curso! Esse é um primeiro passo para iniciar suas análises de saúde utilizando a linguagem R. Ficamos muito felizes que tenha completado todas as aulas!

Para aprender mais sobre o R é importante continuar treinando. Tente aplicar os conhecimentos desse curso em seu cotidiano na vigilância em saúde. Tente incorporar o R aos poucos. No início pode parecer confuso, mas a medida que você treinar se tornará melhor.

Desejamos-lhe sucesso!

## **7. Referências Bibliográficas**

1. Medronho R de A, Bloch KV, Luiz RR, Werneck GL. Epidemiologia. Atheneu; 2009. 685 p.
2. Fávero LP, Belfiore P. Manual de Análise de Dados: Estatística e Modelagem Multivariada com Excel®, SPSS® e Stata®. Elsevier Brasil; 2017. 1832 p.