**Mensurando das mudanças climáticas[[1]](#footnote-1)**

1. **Os objetivos da atividade**

* Utilizar gráficos e medidas sintéticas para discutir a extensão da mudança climática e suas possíveis causas
* Utilizar gráficos de linhas para descrever o comportamento de variáveis do mundo real ao longo do tempo
* Preparar dados em uma tabela de frequência e visualizar distribuições com gráficos de colunas
* Descrever uma distribuição usando média e variância
* Utilizar gráficos de dispersão e o coeficiente de correlação para avaliar o grau de associação entre duas variáveis
* Explicar quais medidas de correlação e as limitações de correlação

1. **Contextualização**

A mudança climática é um dos efeitos do rápido crescimento econômico que ocorreu na maioria dos países desde a Revolução Industrial. É uma questão importante para a formulação de políticas públicas, já que os governos precisam avaliar a gravidade do problema e decidir como mitigá-lo ou mesmo as estratégias para adaptação aos potenciais efeitos das mudanças climáticas, tais como, a ocorrência de eventos extremos, aumento ou redução do volume e intensidade das chuvas e dos períodos de estiagens, aumento do nível do mar, entre outros eventos.

Suponha que você seja um conselheiro político de um país ou de uma região. O governo gostaria de saber mais sobre a extensão da mudança climática e suas possíveis causas. Eles fazem as seguintes perguntas:

* Como podemos saber se a mudança climática está realmente acontecendo ou não?
* Se é real, como podemos medir a extensão da mudança climática e determinar o que está causando isso?

Para responder à primeira questão, devemos observar o comportamento das variáveis ambientais ao longo do tempo para verificar se existem padrões gerais nas condições ambientais que possam ser indicativos da mudança climática. Nesta atividade, nos concentramos em variáveis relacionadas à temperatura.

Para responder à segunda pergunta, podemos examinar o grau de associação entre a temperatura e outra variável, por exemplo, as emissões de CO2 (Dióxido de Carbono), e avaliarmos se existe uma relação plausível entre as duas variáveis, ou se há outras explicações que podem ser observadas.

1. **O comportamento da temperatura média da superfície terrestre ao longo do tempo**

Nesta atividade buscamos examinar com base em dados da NASA as anomalias de temperatura terra-oceano, média do Hemisfério Norte, de 1880 a 2018. A Figura 1 foi preparada com base nestes dados e mostra as temperaturas médias registradas no mundo no período 1880–2018, além de expressar as diferenças da temperatura média de 1951 a 1980. A atividade começa com a preparação de gráficos semelhantes à Figura 1, os quais permitem visualizar os dados e identificar padrões com mais facilidade.

Figura 1 – Anomalias de temperatura terra-oceano, média do Hemisfério Norte: 1880-2018

Fonte: preparado com base em Nasa (https://data.giss.nasa.gov/gistemp/).

Antes de preparar qualquer gráfico, faça o *download* dos dados e certifique-se de entender como a temperatura é medida:

* Acesse a página do *Goddar Institute for Space Studies* da Nasa: <https://data.giss.nasa.gov/gistemp/>
* Faça o download do arquivo “Global-mean monthly, seasonal, and annual means, 1880-present, updated through most recent month” em formato CSV.
* Abra o arquivo e salve o arquivo em formato Microsoft Excel

1. Neste conjunto de dados, a temperatura é medida como "anomalias" em vez de temperatura absoluta. Usando [NASA’s Frequently Asked Questions section](http://tinyco.re/8370199) como referência, explique em suas próprias palavras o que significa "anomalias" de temperatura. Por que os pesquisadores escolheram essa medida específica em detrimento de outras medidas (como a temperatura absoluta)?

Agora prepare alguns gráficos de linhas usando dados mensais, sazonais e anuais, que nos ajudam a procurar padrões gerais ao longo do tempo.

1. Selecione um mês e prepare um gráfico de linhas com a anomalia de temperatura média no eixo vertical e no tempo (1880–2018) no eixo horizontal. Rotule cada eixo apropriadamente e dê ao seu gráfico um título adequado (consulte a Figura 1 como exemplo).
   1. **Desenhando um gráfico de linha de temperatura e tempo**
2. As colunas os títulos "DJF", "MAM", "JJA" e "SON" contêm médias sazonais (médias). Por exemplo, a coluna "MAM" contém a média das colunas de março, abril e maio de cada ano. Prepare um gráfico de linhas separado para cada estação do ano, usando a anomalia de temperatura média para aquela estação do ano no eixo vertical e tempo (1880–2018) no eixo horizontal.
3. A coluna com o título "J-D" contém a anomalia média de temperatura para cada ano.
   1. Prepare um gráfico de linhas com anomalia de temperatura média anual no eixo vertical e tempo (1880–2018) no eixo horizontal. Seu gráfico deve parecer com a figura 1.

**Extensão:** adicione uma linha horizontal que cruze o eixo vertical em 0 e identifique-a como "média 1951-1980".

* 1. O que seus gráficos das questões II a IV (a) sugerem sobre a relação entre temperatura e tempo?
  2. **Preparando um gráfico de linhas e adicionando uma linha horizontal**

1. Agora você tem gráficos para três intervalos de tempo diferentes: mês, estação do ano e ano. Para cada intervalo de tempo, discuta o que podemos aprender sobre os padrões de temperatura ao longo do tempo que talvez não possamos aprender com os gráficos de outros intervalos de tempo.
2. Compare seu gráfico da Questão 4 à Figura 2, que também mostra o comportamento da temperatura ao longo do tempo usando dados retirados da Academia Nacional de Ciências (*National Academy of Sciences*).
   1. Discuta as semelhanças e diferenças entre os gráficos. (Por exemplo, as variáveis dos eixos horizontal e vertical são as mesmas, ou as linhas têm a mesma forma?).
   2. Analisando o comportamento da temperatura ao longo do tempo de 1000 a 1900 na Figura 2, os padrões observados em seu gráfico são incomuns?
   3. Com base nas suas respostas às perguntas IV e V, você acha que o governo deveria se preocupar com as mudanças climáticas?

Figura 2 – Temperaturas do Hemisfério Norte a longo prazo (1000-2006)

Uma imagem contendo captura de tela

Descrição gerada automaticamente

1. **Variação da temperatura ao longo do tempo**

Além das mudanças na temperatura média, o governo também está preocupado que as mudanças climáticas resultem em eventos climáticos extremos mais frequentes. Alguns regiões do Planeta têm sofrido com tempestades mais intensas (fortes) e ondas de calor severas do que no passado, as quais têm causado sérios danos e perturbações na atividade econômica e na sociedade.

O clima se tornará mais extremo e variará mais como resultado da mudança climática? Um artigo do [New York Times](https://www.nytimes.com/interactive/2017/07/28/climate/more-frequent-extreme-summer-heat.html?emc=edit_th_20170729&nl=todaysheadlines&nlid=36015469) ([It’s Not Your Imagination: Summers Are Getting Hotter](https://www.nytimes.com/interactive/2017/07/28/climate/more-frequent-extreme-summer-heat.html?emc=edit_th_20170729&nl=todaysheadlines&nlid=36015469)) usando o mesmo conjunto de dados de temperatura que você está usando para investigar a distribuição das temperaturas e a variabilidade de temperatura ao longo do tempo. Leia o artigo, prestando atenção às descrições das distribuições de temperatura.

Podemos usar a média e a mediana para descrever as distribuições, e podemos usar o [decil](https://pt.wikipedia.org/wiki/Decil) para descrever partes das distribuições. Para visualizar distribuições, podemos usar gráficos de colunas no Excel. Agora, vamos preparar gráficos semelhantes de distribuição de temperatura aos do artigo do New York Times e analisaremos diferentes maneiras de resumir as distribuições.

Para preparar um gráfico de colunas usando os dados de temperatura, primeiro precisamos resumir os dados usando uma tabela de frequência[[2]](#footnote-2). Em vez de usar decis para agrupar os dados, vamos usar intervalos de 0,05, para que as anomalias de temperatura com um valor de -0,3 a -0,025 estejam em um grupo, um valor maior que -0,025 até 0,02 em outro grupo e assim por diante. A tabela de frequências mostra quantos valores pertencem a um grupo específico.

1. Usando os dados mensais de junho, julho e agosto (colunas G a I em sua planilha), prepare duas tabelas de frequência semelhantes à Figura 3 abaixo para os anos 1951 a 1980 e 1981 a 2010, respectivamente. Os valores na primeira coluna devem variar de -0,3 a 1,05, em intervalos de 0,05.

Figura – tabela de frequência (exemplo)

|  |  |
| --- | --- |
| **Range of temperature anomaly (T)** | **Frequency** |
| −0.30 |  |
| −0.25 |  |
| … |  |
| 1.00 |  |
| 1.05 |  |

* 1. **Preparando uma tabela de frequência**

1. Usando as tabelas de frequência da pergunta 1:
   1. Prepare dois gráficos de coluna separados para 1951–1980 e 1981–2010 para mostrar a distribuição das temperaturas, com frequência no eixo vertical e a variação da anomalia de temperatura no eixo horizontal. Seus gráficos devem ser semelhantes aos do artigo do New York Times.
   2. Usando seus gráficos, descreva as semelhanças e diferenças (se houver) entre as distribuições de anomalias de temperatura em 1951–1980 e 1981–2010.

Agora vamos usar nossos dados para analisar diferentes aspectos das distribuições. Primeiro, aprenderemos a usar decilômetros para determinar quais observações são "normais" e "anormais" e aprender a usar a variância[[3]](#footnote-3) para descrever a forma de uma distribuição.

1. O artigo do New York Times considera o terço inferior (o terço mais baixo ou mais frio) de anomalias de temperatura em 1951-1980 como "frio" e o terço superior (o terço mais alto ou mais quente) de anomalias como "quente". Em termos de decil, as temperaturas do 1º ao 3º decil são "frias" e as temperaturas do 7º ao 10º decil ou superior são "quentes" (arredondadas para o decil mais próximo). Use a função PERCENTILE.INC do Excel para determinar quais valores correspondem ao terceiro e sétimo decil, em todos os meses de 1951 a 1980.
   1. **Calculando percentis**
2. Com base nos valores encontrados na pergunta III, conte o número de anomalias consideradas "quentes" em 1981-2010 e expresse isso como uma porcentagem de todas as observações de temperatura nesse período. Sua resposta sugere que estamos experimentando um clima mais quente com mais frequência em 1981-2010? (Lembre-se de que cada decil representa 10% das observações, portanto, 30% das temperaturas foram consideradas "quentes" em 1951-1980).
   1. **Usando a função COUNTIF do Microsoft Excel**
3. O artigo do New York Times discute se as temperaturas se tornaram mais variáveis ao longo do tempo. Uma maneira de medir a variabilidade de temperatura é calcular a variação da distribuição de temperatura para cada estação do ano ("DJF", "MAM", "JJA" e "SON").
4. Calcule a média (média) e a variância separadamente para os seguintes períodos: 1921–1950, 1951–1980 e 1981–2010.
5. Para cada estação do ano, compare as variações em diferentes períodos e explique se a temperatura parece ou não mais variável em períodos posteriores.
   1. **Calculando e entendendo a variância**
6. Usando as descobertas do artigo do New York Times e suas respostas às questões de I a V, discuta se a temperatura parece ser mais variável ao longo do tempo. Você aconselharia o governo a gastar mais dinheiro na mitigação e adaptação aos efeitos de eventos climáticos extremos?
7. **Emissões de carbono e o meio ambiente**

O governo ficou sabendo que as emissões de carbono poderiam ser responsáveis pela mudança climática e pediu que você investigasse se esse é o caso. Para fazer isso, vamos agora analisar as emissões de carbono ao longo do tempo e usar outro tipo de gráfico (gráficos de dispersão) para mostrar sua relação com as anomalias de temperatura. Uma maneira de medir o relacionamento entre duas variáveis é a correlação[[4]](#footnote-4).

Nas perguntas abaixo, vamos preparar gráficos usando os dados de CO2 da Administração Nacional Oceânica e Atmosférica dos EUA. Faça o download da planilha do Excel que contém esses dados ([planilha com dados de CO2](https://tinyco.re/3763425)).

1. Os dados de CO2 foram registrados de um observatório em Mauna Loa, Havaí. Usando um [artigo do Laboratório de Pesquisas do Sistema Terrestre como referência](http://tinyco.re/8193893) (*Earth System Research Laboratory*), explique se você acredita ou não que esses dados são uma representação confiável da atmosfera global.
2. As variáveis "tendência" (*trend*) e "interpoladas" (*interpolated*) são semelhantes, mas não idênticas. Em suas próprias palavras, explique a diferença entre essas duas medidas de níveis de CO2. Por que pode haver variação sazonal nos níveis de CO2?

Agora vamos utilizar um gráfico de linhas para procurar padrões gerais ao longo do tempo.

1. Prepare um gráfico de linhas com níveis de CO2 interpolados e de tendência no eixo vertical e no tempo (a partir de janeiro de 1960) no eixo horizontal. Rotule os eixos e a legenda do gráfico e dê ao seu gráfico um título apropriado. O que este gráfico sugere sobre a relação entre o CO2 e o tempo?

Agora vamos combinar os dados de CO2 com os dados de temperatura e examinar a relação entre essas duas variáveis visualmente, usando gráficos de dispersão e, estatisticamente, usando o coeficiente de correlação[[5]](#footnote-5).

1. Selecione um mês e adicione os dados de tendências de CO2 ao conjunto de dados de temperatura, certificando-se de que os dados correspondam ao ano correto.
2. Prepare um gráfico de dispersão do nível de CO2 no eixo vertical e anomalia de temperatura no eixo horizontal.
3. Calcule e interprete o coeficiente de correlação (Pearson) entre essas duas variáveis.
   1. **Calculando a correlação e preparando um gráfico de dispersão**
4. Escolha dois meses e adicione os dados de tendências de CO2 ao conjunto de dados de temperatura, certificando-se de que os dados correspondam ao ano correto.
5. Prepare um gráfico para cada mês. O que seus gráficos e os coeficientes de correlação sugerem sobre a relação entre os níveis de CO2 e as anomalias de temperatura?
6. Discuta as deficiências do uso desse coeficiente para resumir a relação entre as variáveis.

Mesmo que duas variáveis estejam fortemente correlacionadas entre si, não é necessariamente o caso de que o comportamento de uma variável seja o resultado da outra (uma característica conhecida como causação[[6]](#footnote-6)). As duas variáveis podem estar falsamente correlacionadas. O exemplo a seguir ilustra a correlação espúria[[7]](#footnote-7).

O desempenho acadêmico de uma criança, por exemplo, pode ser correlacionado positivamente com o número de cômodos no tamanho de sua casa ou tamanho de sua casa, mas poderíamos concluir que a construção de um cômodo extra tornaria a criança mais inteligente, ou que um bom desempenho na escola aumentaria sua casa? É mais plausível que renda ou riqueza, que determina o tamanho da casa que uma família pode pagar e os recursos disponíveis para estudar, seja o "fator invisível" nessa relação. Poderíamos também determinar se a renda é a razão para essa correlação espúria comparando as pontuações dos exames para crianças cujos pais têm renda semelhante, mas tamanhos de casas diferentes. Se não houver correlação entre as pontuações do exame e o tamanho da casa, podemos deduzir que o tamanho da casa não estava "causando" pontuações nos exames (ou vice-versa).

1. Considere o exemplo da correlação espúria acima.
   1. Em suas próprias palavras, explique a correlação espúria e a diferença entre correlação e causação.
   2. Dê um exemplo de correlação espúria, semelhante à acima, para níveis de CO2 ou anomalias de temperatura.
   3. Escolha um exemplo de correlação espúria no [site de Tyler Vigen](http://tinyco.re/8861803). Explique se você acredita que é uma coincidência ou se essa correlação pode ser devida a uma ou mais outras variáveis.
   4. Encontre alguns outros exemplos de correlações espúrias nas notícias e explique brevemente por que são espúrias.
   5. **Quais as causas para algumas correlações espúrias?**

O uso de análise estatísticas, como o coeficiente de correlação, pode ajudar a identificar possíveis padrões ou relações entre variáveis, mas não podemos tirar conclusões sobre a causação. Também é importante pensar em outras explicações para o que observamos nos dados e se esperamos que haja uma relação entre as duas variáveis.

No entanto, existem maneiras de determinar se existe uma relação causal entre duas variáveis, por exemplo, observando os processos que conectam as variáveis (como as anomalias de CO2 e temperatura) ou usando um experimento empírico ou natural[[8]](#footnote-8).

1. Preparada com base em Doing Economics: empirical projects (http://www.core-econ.org/doing-economics/index.html). [↑](#footnote-ref-1)
2. Um registro de quantas observações em um conjunto de dados tem um valor específico, um intervalo de valores ou pertence a uma categoria específica. [↑](#footnote-ref-2)
3. A variância é uma medida de dispersão para analisar a distribuição de frequência, sendo igual à média dos quadrados dos desvios da média aritmética da distribuição. A variação é usada para indicar a "dispersão" dos dados. Uma variação maior significa que os dados estão mais dispersos. Exemplo: o conjunto de números 1, 1, 1 tem variância zero (sem variação), enquanto o conjunto de números 1, 1, 999 tem uma alta variância de 2178 (grande dispersão). [↑](#footnote-ref-3)
4. A correlação é uma medida de quão intimamente relacionadas são duas variáveis. Duas variáveis são correlacionadas se o valor de uma variável fornece informações sobre o valor provável da outra, por exemplo, valores altos de uma variável sendo comumente observados juntamente com valores altos da outra variável. A correlação pode ser positiva ou negativa. É negativa quando valores altos de uma variável são observados com valores baixos da outra. A correlação não significa que exista uma relação causal entre as variáveis. Exemplo: quando o tempo está mais quente, as compras de sorvete são mais altas. As vendas de sorvete e a temperatura estão positivamente correlacionadas. Por outro lado, se as compras de bebidas quentes diminuem quando o clima é mais quente, dizemos que as vendas de bebidas quentes e a temperatura estão negativamente correlacionadas. [↑](#footnote-ref-4)
5. O coeficiente de correlação é uma medida numérica de quão intimamente associadas duas variáveis são e se elas tendem a ter valores similares ou diferentes, variando de um valor de 1, indicando que as variáveis tomam valores similares (correlacionados positivamente), a -1, indicando que as variáveis tomam diferentes dissimilares variáveis (correlação negativa ou inversa). Um valor de 1 ou -1 indica que o valor de uma das variáveis permitir-lhe-ia prever perfeitamente o valor da outra. Um valor de 0 indica que uma das variáveis não fornece informações sobre o valor da outra. [↑](#footnote-ref-5)
6. A causação representa uma direção de causa para efeito, estabelecendo que uma mudança em uma variável produz uma mudança em outra. Enquanto uma correlação fornece uma indicação de duas variáveis que se movem juntas (na mesma direção ou em direções opostas), a causalidade significa que existe um mecanismo que explica essa associação. Exemplo: sabemos que níveis mais altos de CO2 na atmosfera levam a um efeito estufa, que aquece a superfície da Terra, portanto, podemos dizer que níveis mais elevados de CO2 são a causa de temperaturas superficiais mais altas. [↑](#footnote-ref-6)
7. A correlação espúria representa uma situação em que uma forte associação linear entre duas variáveis não resulta de qualquer relação direta, mas pode ser devido a uma coincidência ou a outro fator não observado. [↑](#footnote-ref-7)
8. Um estudo empírico explora os controles estatísticos de ocorrência natural, nos quais os pesquisadores não têm a capacidade de designar participantes para grupos de tratamento e controle, como é o caso em experimentos convencionais. Em vez disso, as diferenças na lei, na política, no clima ou em outros eventos podem oferecer a oportunidade de analisar as populações como se tivessem sido parte de um experimento. A validade de tais estudos depende da premissa de que a designação de sujeitos para os grupos de tratamento e controle que ocorrem naturalmente pode ser plausivelmente argumentada como aleatória. [↑](#footnote-ref-8)