

Universidade Federal do ABC (UFABC)
Programa de Pós-graduação em Políticas
Públicas

Professor: Ricardo Ceneviva

20 de março de 2024

Métodos Quantitativos II: Modelos Lineares

1. Antes de construir qualquer gráfico ou tabela, faça o download dos dados e certifique-se de entender como a temperatura é medida:
2. Vá para o site do Goddard Institute for Space Studies (GISS) da NASA em: <https://www.giss.nasa.gov/>
3. Sob o subtítulo '*Combined Land-Surface Air and Sea-Surface Water Temperature Anomalies*', selecione a versão CSV de '*Northern Hemisphere-mean monthly, seasonal, and annual means*' , clique com o botão direito do mouse e selecione '*Salvar link como...*' .
4. O nome padrão deste arquivo é "NH.TS+DSST.CSV". Dê um nome mais adequado – para facilitar sua programação no R e tornar a leitura do seu código mais intuitiva – e salve-o em um local facilmente acessível, como uma pasta na sua área de trabalho ou na sua pasta pessoal.

Exercício 01: Mudanças Climáticas

Nesta base de dados, a temperatura é medida como “*anomalias*” e não como temperatura absoluta. Usando a seção de “*FAQs*” do site do GISS/NASA como referência, explique com suas próprias palavras o que significam “*anomalias*” de temperatura e por que os pesquisadores do GISS/NASA escolheram esta medida específica em detrimento de outras (como a temperatura absoluta)?

As mudanças climáticas decorrentes do rápido crescimento econômico que ocorreu na maioria dos países desde a Revolução Industrial é, talvez, o problema social mais importante do nosso tempo, na medida em que ameaça a sobrevivência da espécie humana no nosso planeta; além de uma questão de Políticas Públicas das mais relevantes, já muitos governos precisam de avaliar com urgência a gravidade do problema e decidir como mitigá-lo.

Suponha que você seja um consultor político de uma pequena nação insular. O governo gostaria de saber mais sobre a extensão das mudanças climáticas e as suas possíveis causas e efeitos. Para tanto, eles a contrataram para responder as seguintes perguntas:

1. Como podemos saber – e mensurar – se as mudanças climáticas estão realmente acontecendo ou não?
2. Se as mudanças climáticas são reais, como podemos medir a extensão dessas alterações e tentar determinar se suas causas estão, de fato, ligadas à atividade econômica do ser humano no nosso planeta?

Para responder à primeira questão, devemos examinar o comportamento das variáveis ambientais ao longo do tempo para verificar se existem padrões gerais nas condições ambientais que possam ser indicativos de mudanças climáticas. Por hora, nos concentramos nas variáveis relacionadas à temperatura.

Para responder à segunda questão, devemos examinar o grau de associação entre a temperatura e outra variável, os níveis de CO₂ na atmosfera. Adicionalmente, devemos considerar se existe uma relação plausível entre as duas variáveis e tentar mensurar a magnitude da associação entre elas; ou, alternativamente, se existem outras explicações para o que observamos.

Roteiro para responder às perguntas com base nos dados

1. Primeiro, você deve importar a base de dados para o **R**
2. Agora, crie um gráfico de linhas – uma série histórica de temperaturas – usando dados mensais, sazonais ou anuais, que nos ajudam a procurar padrões gerais ao longo do tempo. (DICA: Escolha um mês e trace um gráfico de linhas com 'anomalia de temperatura média' no eixo vertical (eixo Y) e tempo – de 1880 até o último ano disponível – no eixo horizontal (eixo X). Rotule cada eixo apropriadamente e dê ao seu gráfico um título adequado.
3. Na base de dados, as colunas denominadas DJF, MAM, JJA e SON contêm médias sazonais. Por exemplo, a coluna MAM contém a média das colunas de março, abril e maio de cada ano. Trace um gráfico de linhas separado para cada estação, usando a anomalia de temperatura média para essa estação no eixo vertical e o tempo (de 1880 até o último ano disponível) no eixo horizontal.
4. A coluna denominada J-D contém a anomalia de temperatura média para cada ano. (1) Trace um gráfico de linhas com anomalia de temperatura média anual no eixo vertical e tempo (de 1880 até o último ano disponível)

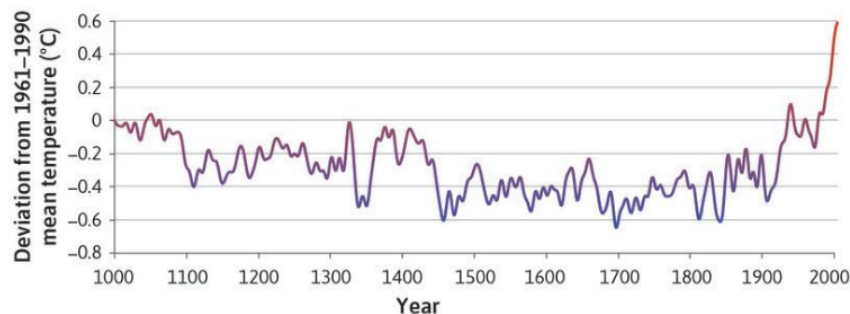


Figura 1: Northern hemisphere temperatures over the long run (1000–2006).

no eixo horizontal. Agora, (2) adicione uma linha horizontal que cruze o eixo vertical em 0 e rotule-a como ‘média de 1951–1980’.

5. Na sua opinião, o que os gráficos das questões 2 a 4 sugerem sobre a relação entre temperatura (Y) e tempo (X)?
6. Agora você tem gráficos para três intervalos de tempo diferentes: mês (Pergunta 2), estação do ano (Pergunta 3) e ano (Pergunta 4). Para cada intervalo de tempo, discuta o que podemos aprender especificamente sobre os padrões de temperatura ao longo de cada medida de tempo, que possivelmente não fiquem tão claros nas outras medidas de tempo.
7. Compare o gráfico da Pergunta 4 com a Figura 1, apresentada acima, que também mostra o comportamento da temperatura ao longo do tempo usando dados retirados da Academia Nacional de Ciências dos EUA.
 - (A) Discuta as semelhanças e diferenças entre os gráficos. (Por exemplo, as variáveis dos eixos horizontal e vertical são iguais ou as linhas têm o mesmo formato?)
 - (B) Observando o comportamento da temperatura ao longo do tempo, de 1000 a 1900 na Figura 1, os padrões observados em seu gráfico são incomuns?
 - (C) Com base nas suas respostas anteriores, na sua opinião, o que você acredita que governo insular, que lhe contratou para fazer essa análise, tem razões reais (isto é, baseadas em evidências) para se preocupar com as mudanças climáticas?

Exercício 02: Eventos Climáticos Extremos

Além das mudanças na temperatura média, o governo insular, que lhe contratou, também está preocupado com a possibilidade de as mudanças climáticas causem

eventos climáticos extremos mais frequentes. A ilha sofreu algumas grandes tempestades e fortes ondas de calor no passado, que causaram graves danos e perturbações na atividade econômica.

As mudanças climáticas ocasionarão eventos climáticos extremos como ondas de calor e tempestades? Um artigo do *New York Times* (disponível aqui) usa o mesmo conjunto de dados de temperatura que você tem usado para investigar a distribuição de temperaturas e a variabilidade de temperatura ao longo do tempo. Primeiro, leia o artigo, prestando muita atenção às descrições das distribuições de temperatura.

Nós podemos usar a média e a mediana para descrever distribuições empíricas e podemos usar decis e quantis para descrever partes das distribuições. Para visualizar as distribuições, podemos usar histogramas de frequência. Nesse exercício, vamos criar gráficos de distribuição de temperatura semelhantes aos do artigo do *New York Times* e examinar diferentes maneiras de resumir as distribuições.

Para criar um histograma de densidades usando os dados de temperatura que temos, primeiro precisamos resumir os dados usando uma tabela de frequência. Em vez de usar decis para agrupar os dados, usaremos intervalos de 0,05, de modo que anomalias de temperatura com valor de $-0,3$ a $-0,25$ estarão em um grupo, valores maiores que $-0,25$, até $0,2$, em outro grupo, e assim por diante. A tabela de frequência mostrará quantos valores pertencem a cada um dos 20 grupos de temperatura.

1. Utilizando os dados mensais para os meses de junho, julho e agosto, crie duas tabelas de frequência, para os anos 1951–1980 e 1981–2010, respectivamente. Os valores da primeira coluna devem variar de 0,3 a 1,05, em intervalos de 0,05.
2. Usando as tabelas de frequência construídas no passo anterior, responda às seguintes perguntas:
3. Construa dois histogramas. um para os anos 1951–1980 e outro para 1981–2010, exibindo a distribuição de temperaturas, com as frequências no eixo vertical (Y) e a faixa de anomalia de temperatura no eixo horizontal (X). Seus gráficos devem ser mais ou menos parecidos aos do artigo do *New York Times*.
4. Com base nos gráficos criados no passo anterior, descreva as semelhanças e diferenças (se houver) entre as distribuições das anomalias de temperatura em 1951–1980 e 1981–2010.
5. Nesse passo, usaremos nossos dados para examinar diferentes aspectos das distribuições de temperatura. Primeiramente, iremos usar decis para determinar quais observações são “normais” e “anormais” e, em seguida, aprenderemos como usar a variância para descrever a forma de uma distribuição. Então vamos lá! O artigo do *New York Times* considera o terço inferior (o terço mais baixo ou mais frio) das anomalias de temperatura

em 1951-1980 como “frio” e o terço superior (o terço mais alto ou mais quente) das anomalias como “quente”. Em termos de decil, as temperaturas no 1º ao 3º decil são “frias” e as temperaturas no 7º ao 10º decil ou acima são “quentes” (arredondadas para o decil mais próximo). Use a função ***quantile*** de R para determinar quais valores correspondem ao 3º e 7º decil em todos os meses de 1951–1980.

6. Com base nos valores encontrados no passo anterior, compute o número de anomalias consideradas “quentes” em 1981–2010 e expresse-o como uma percentagem de todas as observações de temperatura nesse período. A sua resposta sugere que experimentamos um clima mais quente com mais ondas de calor no período de 1981–2010? (Lembre-se de que cada decil representa 10% das observações, portanto 30% das temperaturas foram consideradas “quentes” em 1951-1980.)
7. O artigo do *New York Times* discute se as variações de temperatura se tornaram mais intensas ao longo do tempo. Uma maneira de medir a variabilidade da temperatura é calculando a variância da distribuição de temperatura. Então, para cada estação do ano (DJF, MAM, JJA e SON): (1) Calcule a média, a mediana e a variância separadamente para os seguintes períodos: 1921–1950, 1951–1980 e 1981–2010. (2) Para cada estação do ano, compare as variações em diferentes períodos e explique se as variações de temperatura parece ou não ser mais intensas nos períodos posteriores.

Exercício 03: Emissões de Carbono e Mudanças Climáticas

O governo da pequena ilha que lhe contratou ouviu dizer que as emissões de carbono podem ser responsáveis pelas alterações climáticas e pediu-lhe que investigasse se este é, de fato, o caso. Para tanto, vamos agora analisar as emissões de carbono ao longo do tempo e utilizar outro tipo de gráfico, o gráfico de dispersão, para mostrar a sua relação com as anomalias de temperatura. Uma forma de medir a relação entre duas variáveis numéricas por meio da correlação ou, mais especificamente, por meio do índice de correlação linear de Pierson.

Nas questões abaixo, faremos gráficos utilizando os dados de emissões atmosféricas de CO₂ da *National Oceanic and Atmospheric Administration* dos EUA. Em anexo, segue uma planilha Excel contendo esses dados. Salve os dados como um arquivo csv e importe-os para R.

1. Os dados de emissões atmosféricas de CO₂ foram coletados em um observatório em *Mauna Loa*. Usando o artigo do Earth System Research Laboratory, disponível nesse link:

<https://gml.noaa.gov/ccgg/about/co2measurements.html>

como referência, explique se, na sua opinião, esses dados são uma representação confiável da atmosfera global e justifique sua resposta.

2. As variáveis *trend* e *interpolated* são semelhantes, mas não idênticas. Com as suas próprias palavras, explique a diferença entre estas duas medidas dos níveis de CO₂. Por que pode haver variação sazonal nos níveis de CO₂ na atmosfera?
3. Agora, usaremos um gráfico de linha (i.e. uma série temporal) para procurar padrões gerais ao longo do tempo. Portanto, construa um gráfico de linhas com níveis de CO₂ *trend* e *interpolated* no eixo vertical (eixo Y) e tempo (a partir de janeiro de 1960) no eixo horizontal (eixo X). Rotule os eixos e coloque uma legenda do gráfico e dê ao seu gráfico um título apropriado. Agora, responda: o que este gráfico sugere sobre a relação entre os níveis de CO₂ ao longo do tempo?
4. Nesse passo, combinaremos os dados de CO₂ com os dados de temperatura da exercício 1 e, em seguida, examinaremos a relação entre essas duas variáveis visualmente usando gráficos de dispersão e estatisticamente usando o coeficiente de correlação. Para tanto, escolha um mês e adicione os dados de *trend* de CO₂ ao conjunto de dados de temperatura do exercício 1, certifique-se de que os dados correspondem ao ano correto. (1) Faça um gráfico de dispersão do nível de CO₂ no eixo vertical (eixo Y) e da anomalia de temperatura no eixo horizontal (eixo X). (2) Calcule e interprete o coeficiente de correlação (Pearson) entre essas duas variáveis. (3) Discuta as deficiências do uso deste coeficiente para resumir a relação entre variáveis.
5. Agora, refaça o gráfico de dispersão; porém, inverta as variáveis de eixo, colocando anomalia de temperatura no eixo vertical (eixo Y) e níveis de CO₂ no eixo horizontal (eixo x). Calcule e interprete o coeficiente de correlação (Pearson) entre essas duas variáveis e compare os resultados com aqueles encontrados no passo anterior
6. Agora, usando a temperatura como variável dependente, faça a regressão linear da temperatura nos níveis de CO₂ na atmosfera. (1) Construa a equação matemática que expressa esse modelo linear. (2) Interprete os resultados e explique as os coeficientes estimados, os erros-padrão (SE), a estatística t e os intervalos de confiança (IC).

Por favor, não reproduzir ou citar sem o consentimento expresso do autor, atualizada em: 20 de março de 2024.