

**Nome do aluno:** Guilherme Palazzo

**Número matrícula ITA:** 101127

**Número matrícula UNIFESP:** 160095

**Assunto:** Exercício no R de Cointegração - Lista 6

A análise da série será realizada de acordo com os passos abaixo:

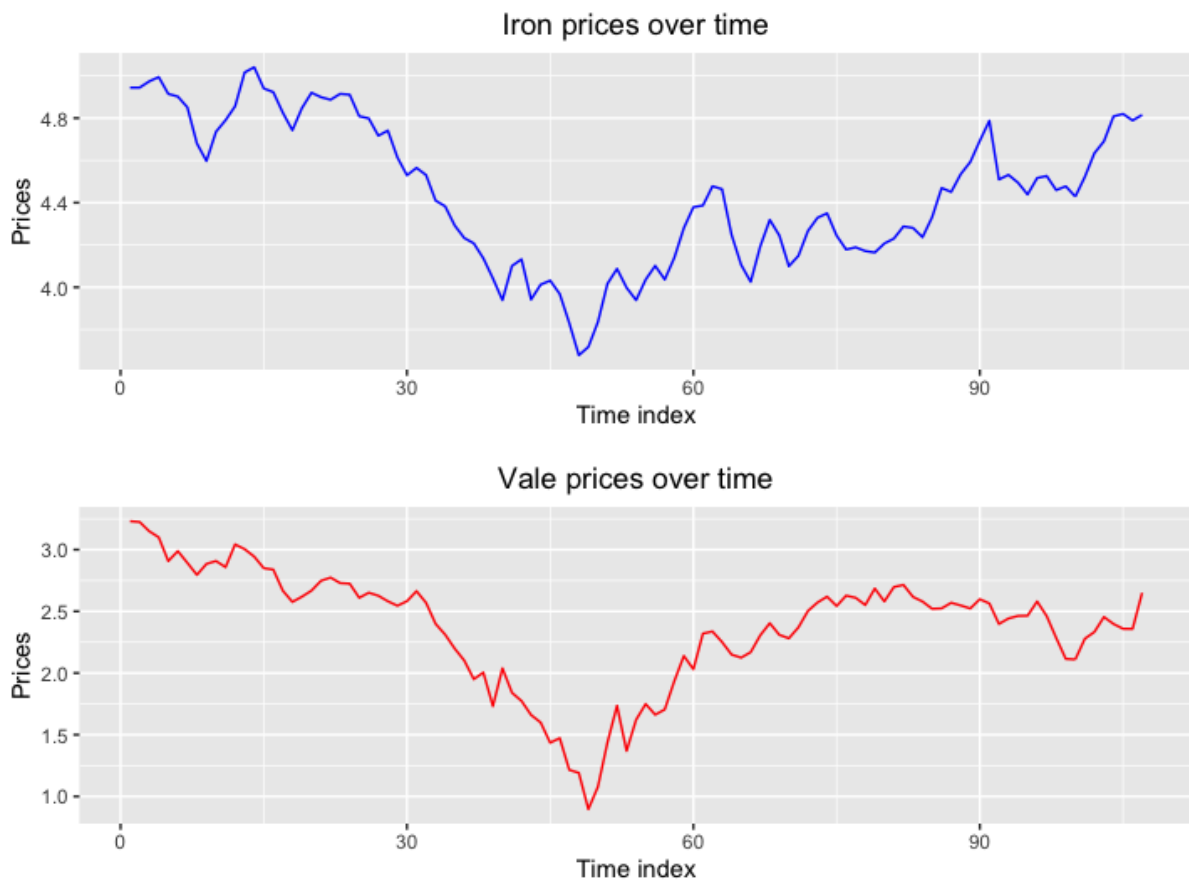
- Plot das séries para análise visual inicial
  - valores das séries ao longo do tempo
  - histograma
- Exploração inicial
  - estatística descritiva para entender inicialmente as séries de um ponto de vista numérico
  - missing: verificação de necessidade de input de dados nulos

Após essas análises iniciais, serão respondidas as perguntas:

- As séries são estacionárias  $I(0)$ ? Justifique sua resposta.
- Ao nível de significância de 5%, a vale causa ferro ou vice-versa ou nada?
- Qual será a variação (%) esperada na vale em  $t+1$ , se o residuo da sua regressão com o minério de ferro em  $t$  for igual a 10%.

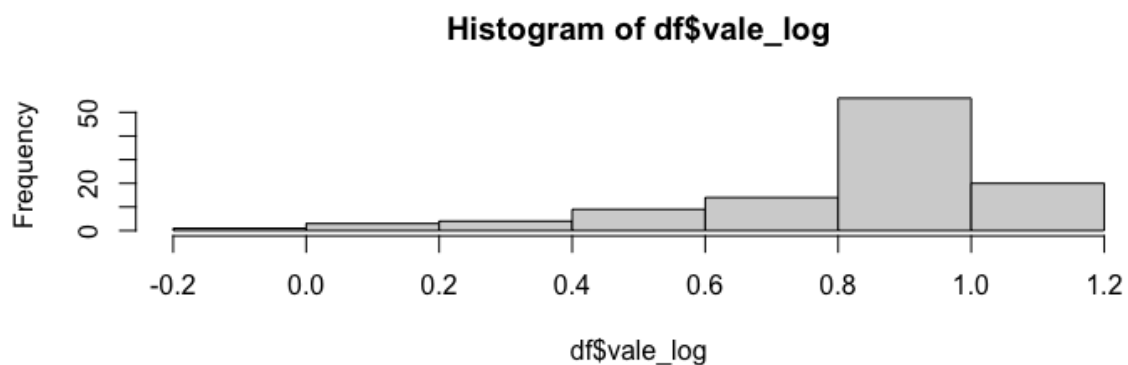
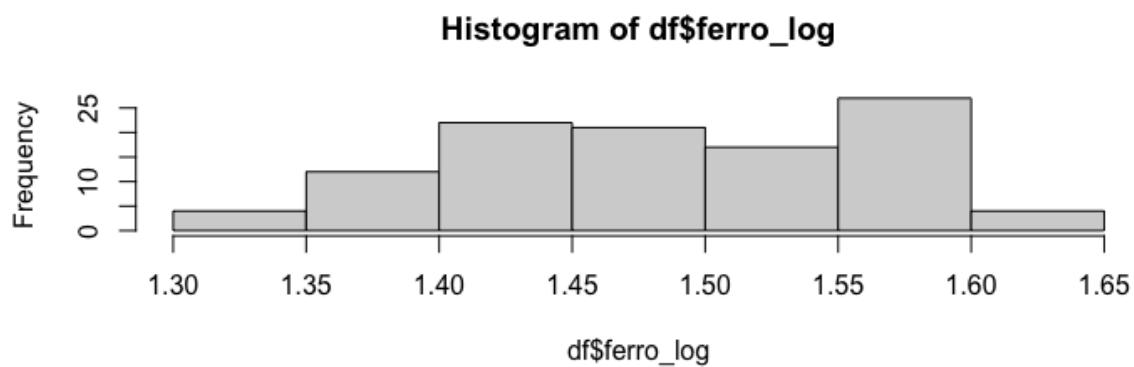
## Plots

Line plots: visualização de tendência



Uma análise rápida inicial nos sugere que as séries não são estacionárias.

Histograma: distribuição do dado



### Exploração inicial

Métrica	Vale	Ferro
Qtd Valores	107	107
Qtd Nulos	0	0
Min	0.896	3.678
Max	3.231	5.041
Mediana	2.504	4.451
Média	2.366	4.443
Variância	0.230	0.116
Desv Pad	0.480	0.340

Sanity check: não há valores negativos e nem nulos.

### Perguntas da Lista

**As séries são estacionárias I(0)? Justifique sua resposta.**

Conforme visto em aula e pesquisando na internet, será utilizado o teste de **Dickey-Fuller**

**Aumentado (ADF)** para responder a pergunta. A hipótese nula nos diz que há raiz unitária, enquanto a alternativa diz que não há.

A quantidade de atrasos (lags) foi definida utilizando-se a metodologia padrão do método **adf.test** da biblioteca **tseries** do R: **trunc((length(x)-1)^(1/3))**. Isso nos dá o valor de 4 lags.

Utilizando o ADF com 4 lags, obtém-se os seguintes valores para as séries de dados após normalização pela aplicação de log:

- vale\_log
  - valor teste ADF = -1.7711
  - p-value = 0.6716
- ferro\_log
  - valor teste ADF = -0.88543
  - p-value = 0.9514

A interpretação dos valores é: quanto mais negativo os valores do ADF, maior a chance de rejeitar a hipótese de existência de raiz unitária. Ou seja, quanto mais negativo, maior a chance de ser estacionário. Como os valores são próximos a 0 e os p-value são maiores que um nível de significância de 5%, falha-se em rejeitar a hipótese nula. Sendo assim, conclui-se que as séries são não estacionárias.

#### **Ao nível de significância de 5%, a vale causa ferro ou vice-versa ou nada?**

Conforme visto em aula e pesquisando na internet, será utilizado o teste de **causalidade de Granger** para determinar se uma das séries é útil para prever a outra. O teste possui como hipótese nula que a série X não causa a série Y, e hipótese alternativa que X causa Y. Ou seja, considerando um certo lag da série X, ela tem poder de prever o valor de Y em um período futuro.

No exercício anterior, verificamos que as séries **vale\_log** e **ferro\_log** são não estacionárias. Portanto, deve-se tirar a primeira diferença delas antes de calcular o teste de Granger, visto que esse teste assume que as séries são estacionárias.

Por último, sabe-se que o teste de Granger é sensível a quantidade de lags utilizados. Dito isso, a quantidade de lags será definida utilizando o critério AIC (Critério de Informação de Akaike) com o máximo de 12 lags possíveis em um modelo VAR.

Aplicando-se o método **causality** da biblioteca **vars** com as características mencionadas, tem-se o seguinte resultado:

- p-value da vale causando ferro: 0.0005006
- p-value do ferro causando vale: 0.1565

Analisando o resultado, pode-se afirmar que vale causa ferro com uma significância estatística menor que 5%.

#### **Qual será a variação (%) esperada na vale em t+1, se o residuo da sua regressão com o minério de ferro em t for igual a 10%.**

Considerando os 60 primeiros índices de tempo do dataframe (de 1 a 60), temos um resíduo de 10,4%. Dessa forma, utilizarei esses 60 valores para estimar o modelo de regressão e a previsão de valor será feita utilizando o índice de tempo 61 (t61). Esse valor de previsão será comparado com o valor real obtido em t61 e então será dada a variação percentual entre o que realmente aconteceu e a previsão do modelo.

Valor real da vale em t61 = 0.8417504

Valor predito da vale em t61 = 0.7472456

Considerando o valor predito como precedessor do valor real para cálculo da variação, o delta percentual entre predito e real é de aproximadamente 12,65%.

**Código em R utilizado para reproduzir os plots e gerar os resultados obtidos**

[https://drive.google.com/file/d/1-3SIO1qW2lv\\_SvQaQkE-ABb9UaSGxFJV/view?usp=sharing](https://drive.google.com/file/d/1-3SIO1qW2lv_SvQaQkE-ABb9UaSGxFJV/view?usp=sharing)